

Latin American Women's Rights Movements: Tracing Online Presence through Language, Time and Space

Archives Unleashed 2023 Cohort Meeting

VANCOUVER, 22 JUNIO 2023

CC BY

Archives Unleashed + ARCHIVE-IT

VANCOUVER, 22 JUNE 2023

Team



Sylvia Fernández Quintanilla

University of Texas at San Antonio



Rosario Rogel Salazar

Universidad Autónoma del Estado de México



Alan Colín Arce

Universidad Autónoma del Estado de México



Abraham García Monroy

Universidad Autónoma del Estado de México



Verónica Benítez Pérez

Universidad Autónoma del Estado de México



Hejin Shin

University of Texas at San Antonio



VANCOUVER, 22 JUNE 2023

Expectation

Develop a historical analysis of the websites from women's rights movements in Mexico and Latin America.

Study how these movements relate to women's rights movements across the world.

Reality

Lack of web archives with content in Spanish and created in Latin America.

Lack of infrastructure for processing large amounts of data.

Web archives blur national boundaries.

Our updated research questions

1. What are the national differences and similarities in the issues addressed by human rights and feminist organizations in Latin America?

2. Do Spanish and English human rights reports on Latin America address the same issues?

3. What are the most common places, organizations, and people mentioned in Latin American human rights reports?

Expectation

- Human Rights (Columbia University Libraries)
- Issues in Women's Rights (McGill University)
- Sophia Smith Collection: Women's History Archive (Smith College)
- Resistance (Columbia University Libraries)
- #metoo collections (Schlesinger Library)

Reality

- Human Rights (Columbia University Libraries)
- Activismos feministas en América Latina (Huellas Incómodas)



Reframing our research approach

1. Creating our own web archive: addressing and questioning a gap in the digital cultural record.

2. Extracting a subset of the Human Rights collection

3. Analyzing pdfs instead of the full text of websites: human rights reports.





VANCOUVER, 22 JUNE 2023

Why create our own collection?

Archives Unleashed + ARCHIVE-IT

VANCOUVER, 22 JUNE 2023

Web Archiving Initiatives



Second-level digital divide
(Hargittai & Hsieh, 2013)

North / South divide

PgomesFCCN, Map of Web archiving initiatives.
CC BY-SA 4.0 via Wikimedia Commons



(Colin-Arce et al., 2023)

Web Archiving Initiatives in Latin America

Only six web archives created by Latin American institutions since 2016 in 3 countries and 1 US territory.

Web archiving is not yet part of the work and research agendas for most libraries and universities in the region.

Practical reasons for creating our own collection

1 Language

Selecting only Spanish websites with a fairer geographic coverage.

2 Space

Smaller selection of websites, but they are all significant for our research interests.

3 Diversity

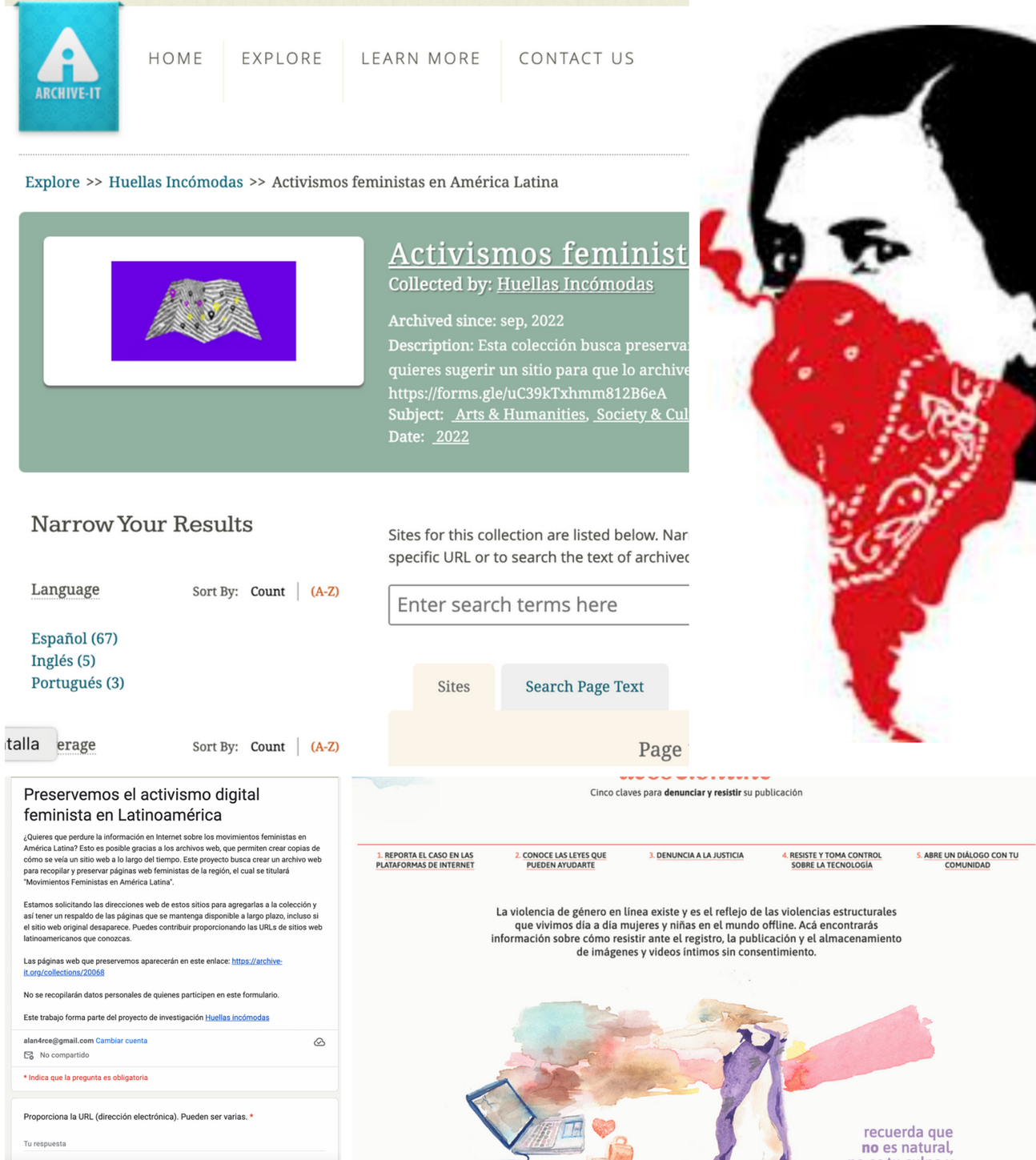
Cover a broader range of feminist issues.

Archives Unleashed + ARCHIVE-IT

Our experience creating a collection

- Better understanding of web archiving
- Better overview of digital feminist activism in the region

- Time consuming
- Need for appraisal, selection and evaluation policies



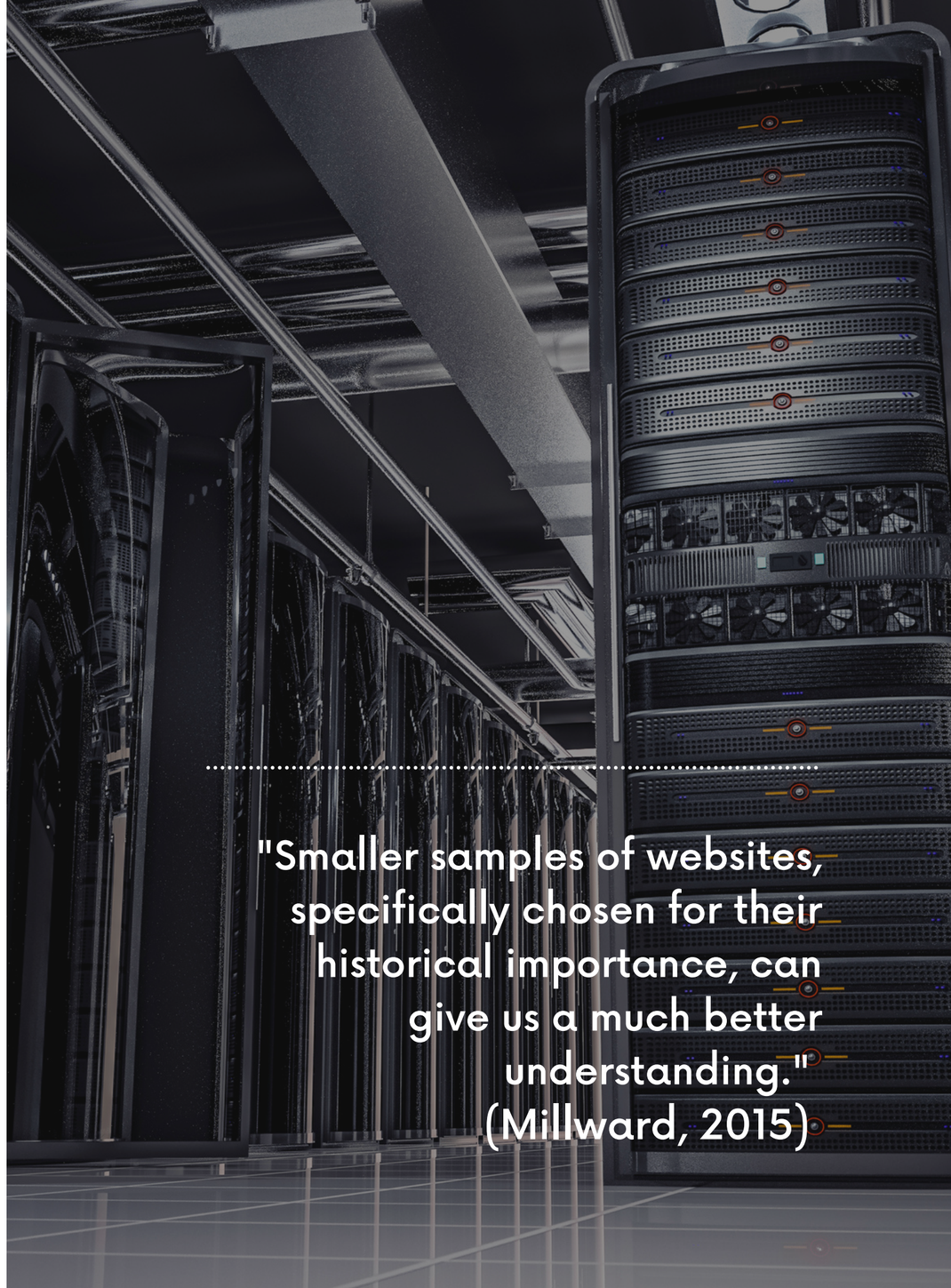


VANCOUVER, 22 JUNE 2023

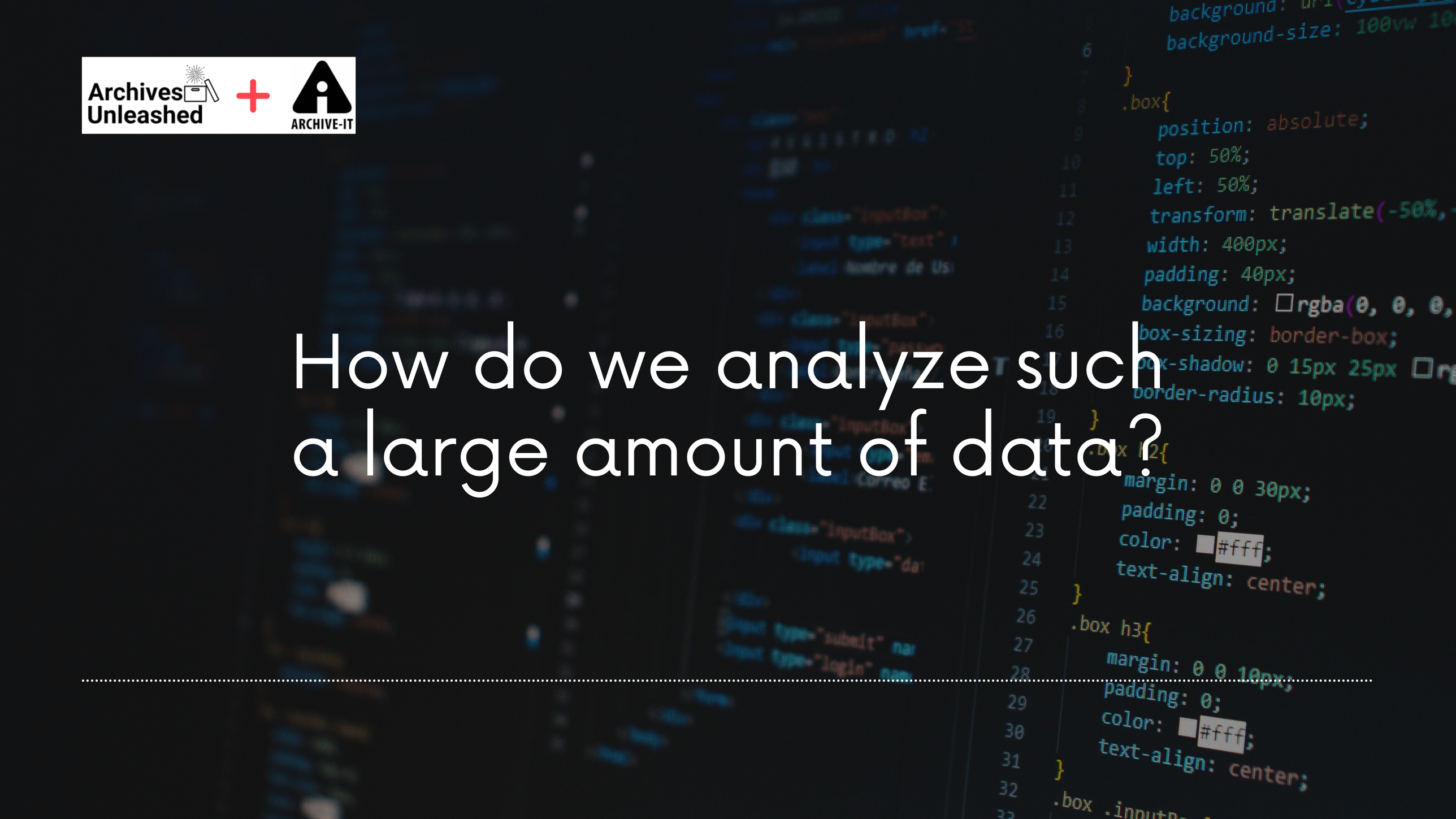
Extracting a subset of the Human Rights collection

Human Rights collection

- 19.5 TB of data
- 1,208 archived websites
- +2 million archived pdfs
- 77 archived websites on women's rights worldwide



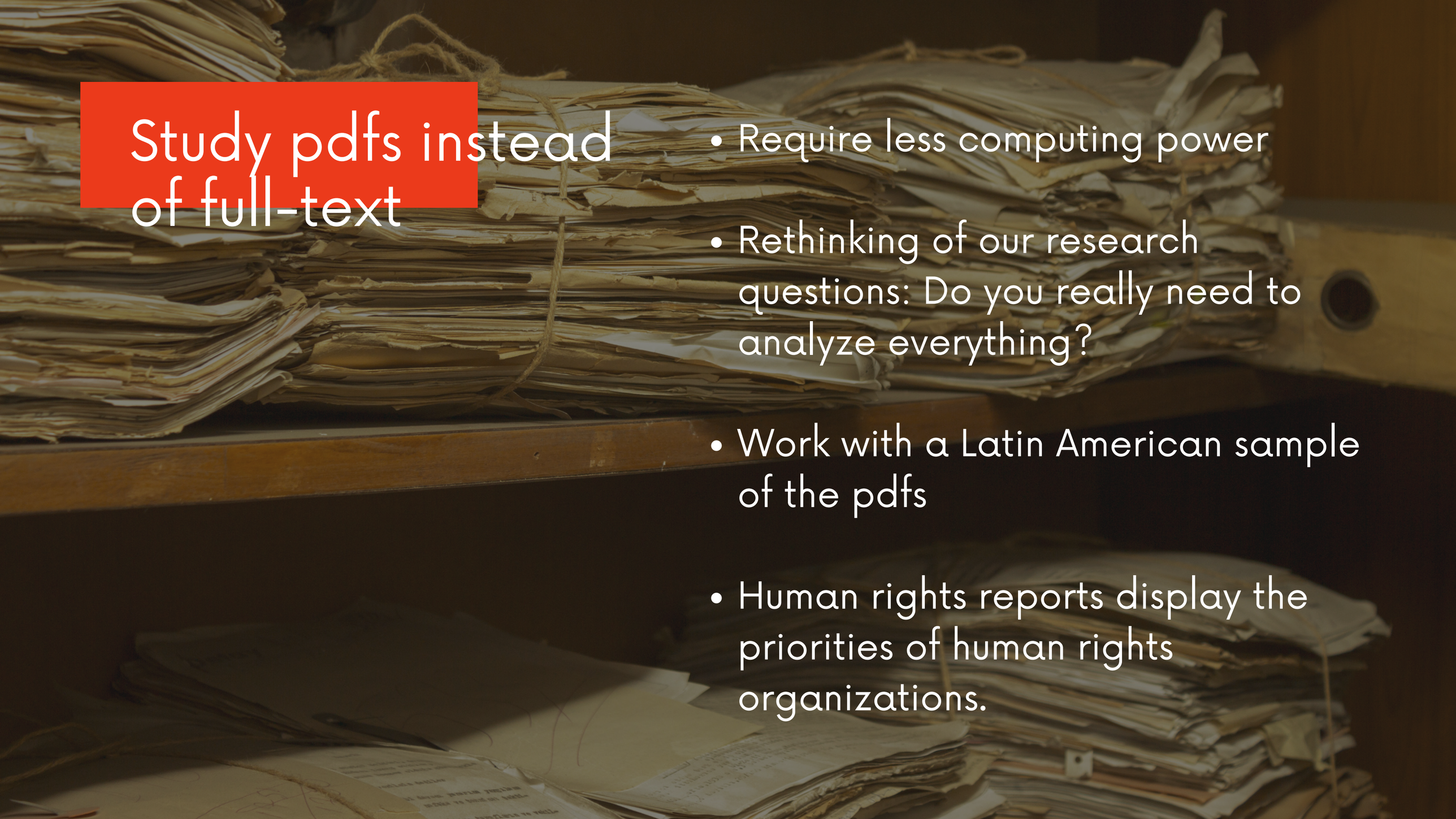
"Smaller samples of websites, specifically chosen for their historical importance, can give us a much better understanding."
(Millward, 2015)


Archives Unleashed
+
ARCHIVE-IT
How do we analyze such a large amount of data?



> Having the right data is usually better than having more data (Borgman, 2015).

> Addressing the crisis in web archive use can be done through the creation of thematic or issue based-collections from existing archives (Rogers, 2019).



Study pdfs instead of full-text

- Require less computing power
- Rethinking of our research questions: Do you really need to analyze everything?
- Work with a Latin American sample of the pdfs
- Human rights reports display the priorities of human rights organizations.



“

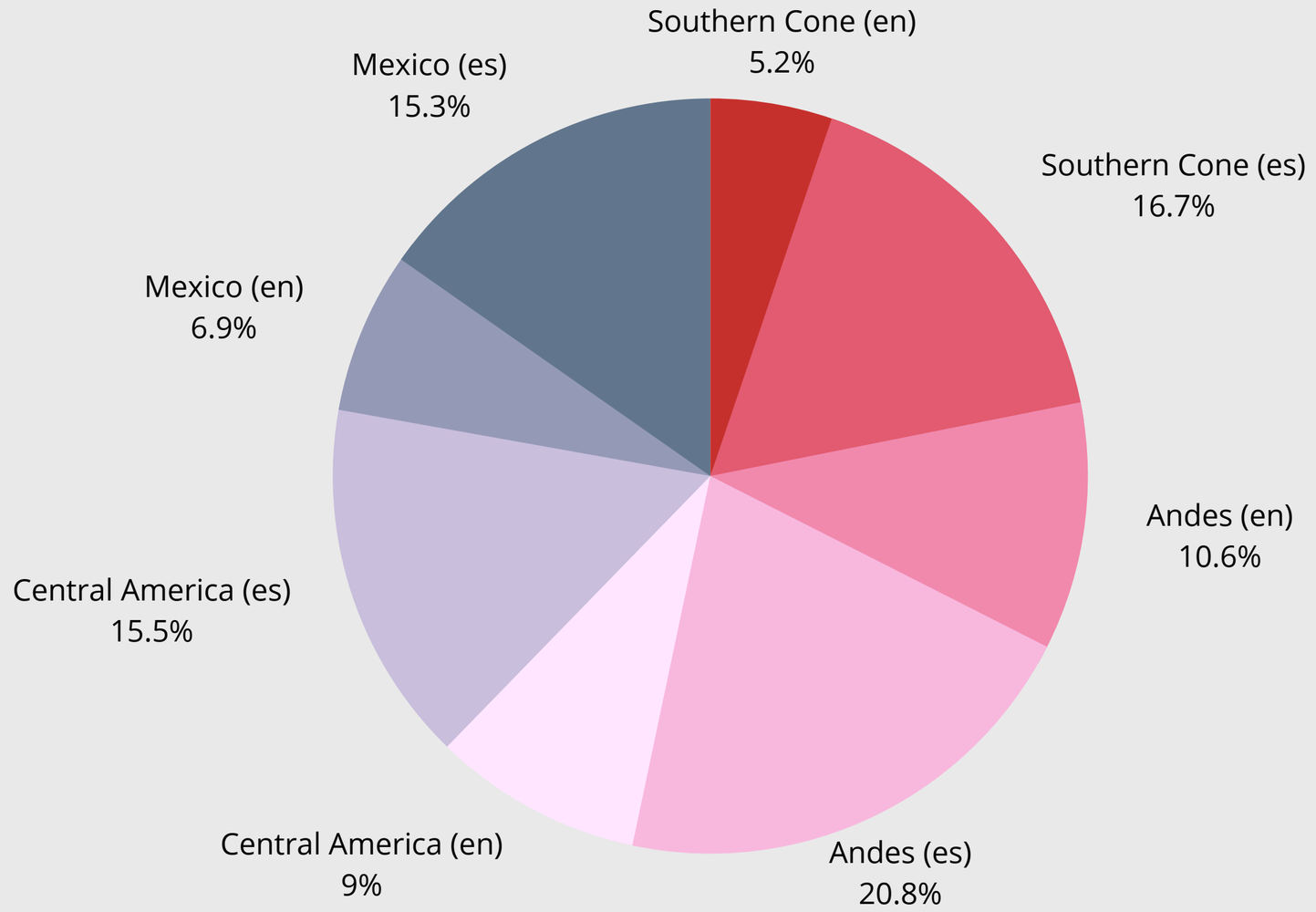
It’s important to think through the kinds of archives we are leaving for those who come after us, and if that archive cannot be read, cannot be seen, cannot be processed because it is simply too large and undifferentiated then we risk burying our subjects in a new way.

(Wernimont, 2013)

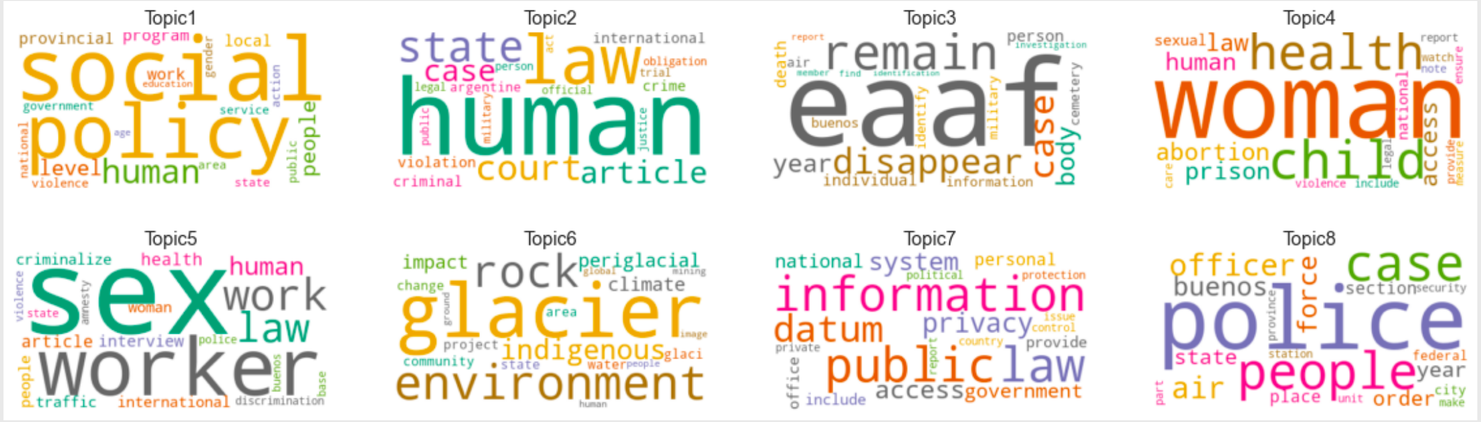
Results

Data

9,923 pdfs
6,777 in Spanish
3,146 in English
16 countries



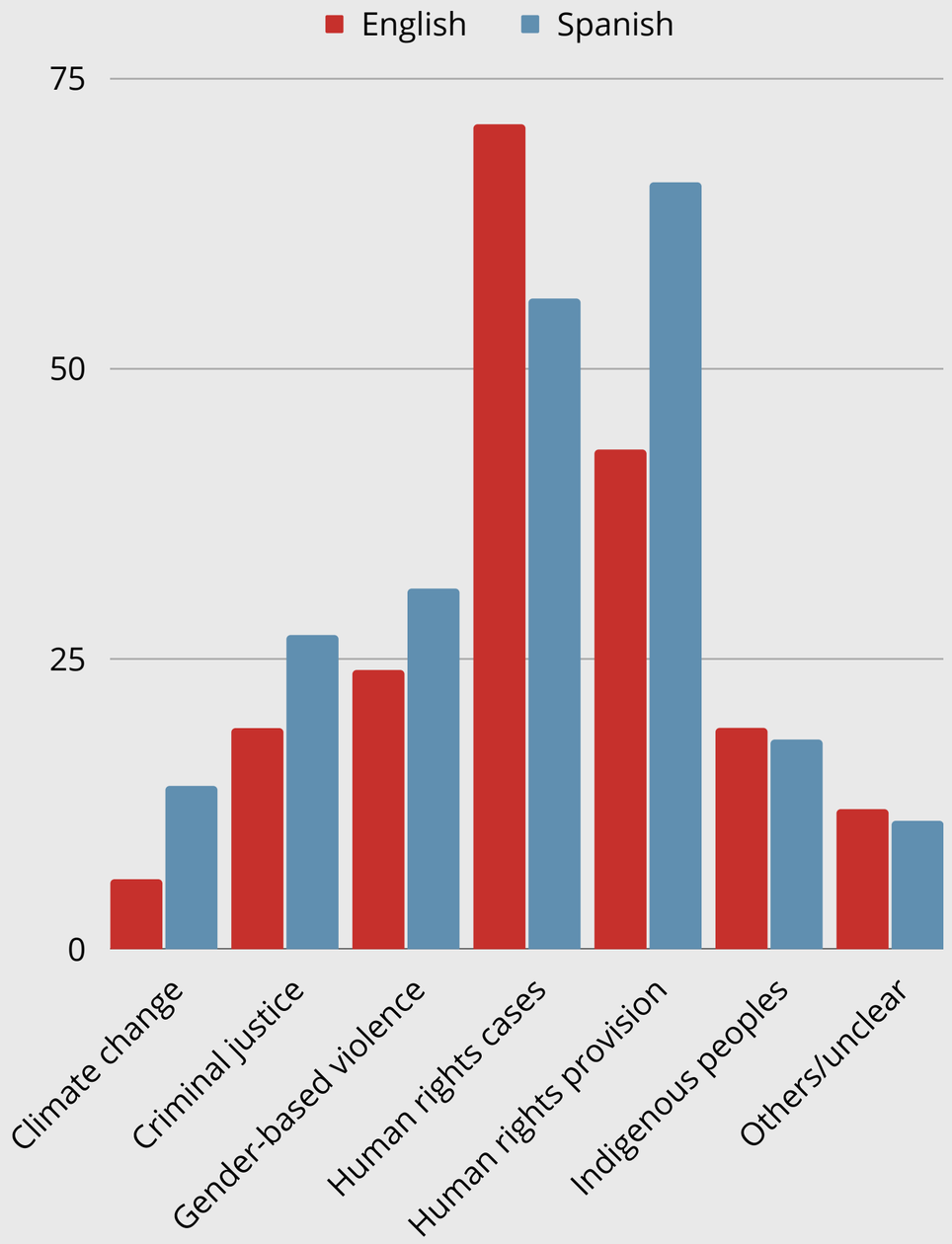




Bilingual topic model (e.g. Argentina)

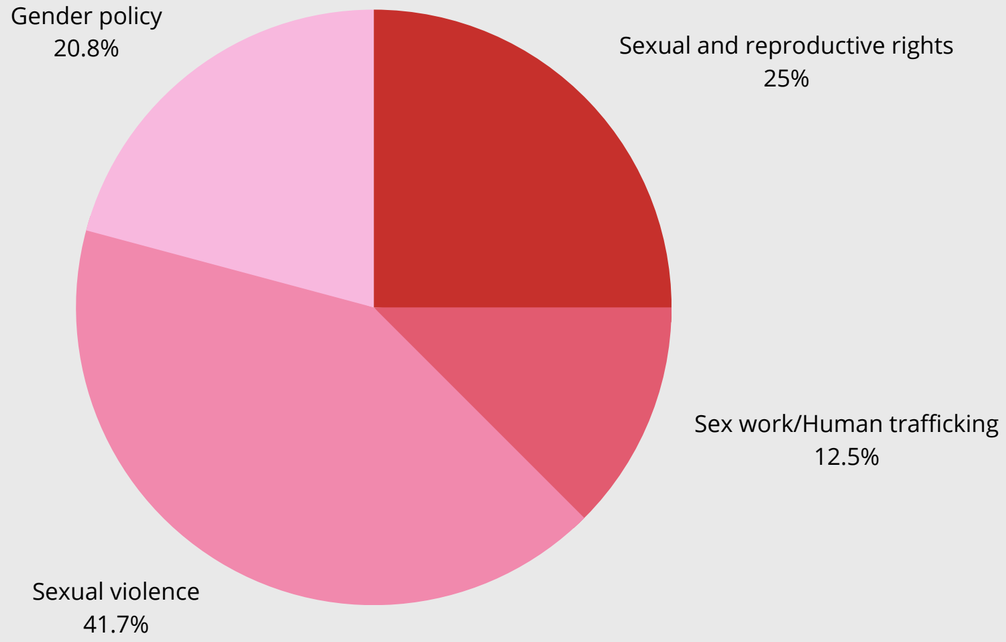
Topic labels

- Seven topic labels after topic modeling.
- Human Rights cases are more frequent in English.
- Human Rights provision is more frequent in Spanish

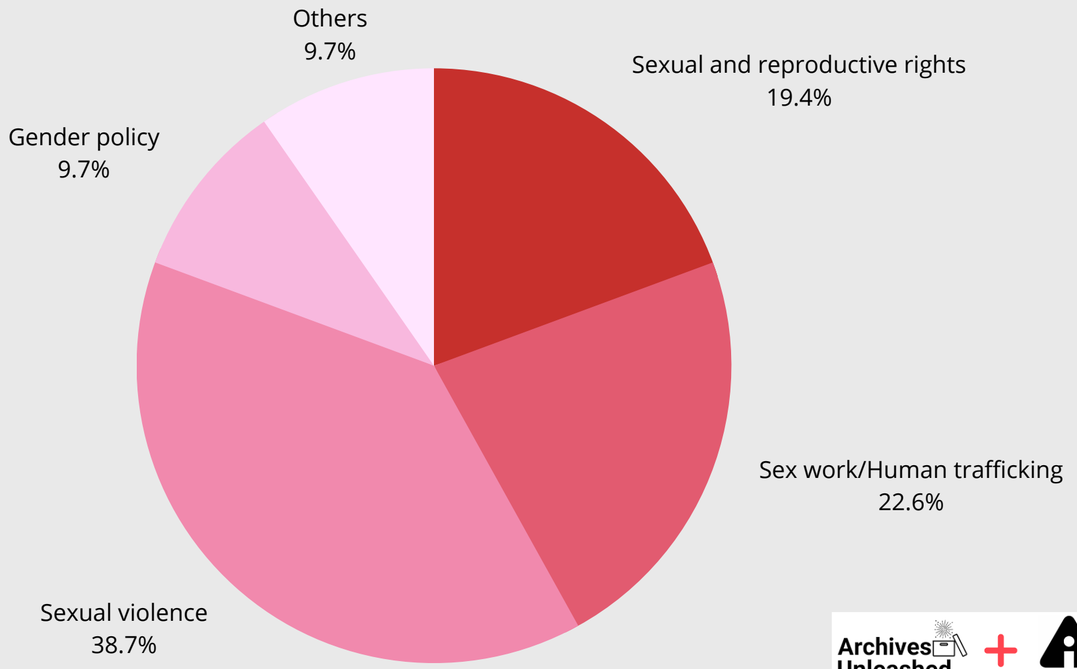


Zooming in the gender-based violence topic

English



Spanish



Archives Unleashed + ARCHIVE-IT

Insights

Diversity

Large scale analysis of gender-based violence in Latin America should be careful when generalizing results, given how gender issues in the region are heterogeneous.

Intersectionality

The frequency of indigenous issues on the reports indicates the need for an intersectional analysis of gender-based violence.

Archival gaps

Important feminist issues are absent in these reports, such as domestic labor, lesbian and transgender issues, or migration.



Archives Unleashed + ARCHIVE-IT

VANCOUVER, 22 JUNE 2023

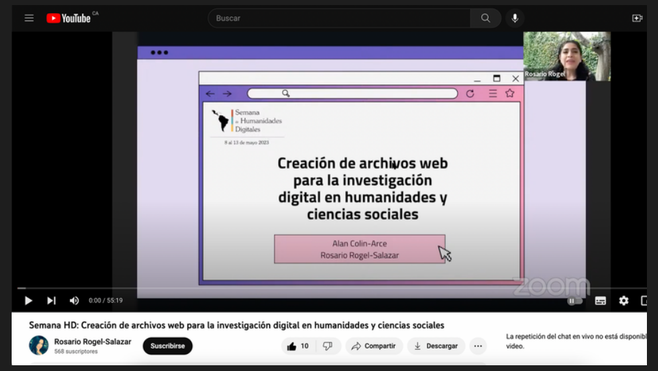
Fostering web archiving in Latin America

Archives Unleashed + ARCHIVE-IT

VANCOUVER, 22 JUNE 2023

Fostering web archiving in Latin America

Speaking in regional and international conferences



Spanish Wikipedia entry for web archiving



Local workshops





VANCOUVER, 22 JUNE 2023

We want to hear from you!

How can we promote web archiving in Latin America?

Future steps

1

Perform text analysis on the full text and pdfs of our collection

2

Continue with web archiving presentations in Latin America

3

Create web archiving resources in Spanish

4

Share our data openly with the information of the 13,000 Latin American pdfs



Discussions of web archives have focused mostly on technical issues, while the politics of web archives have been largely unexplored.

Question the abundance

1

Do all regions in the world have the same amount of digital/digitized historical sources available online?

2

Who is creating and maintaining this abundance of sources?

3

Are the barriers of entry to web archiving preventing women and people of color from getting started with web archives?

Final remarks

1

Whose stories are stored in web archives? What languages, countries and groups are (over)represented?

2

There are more countries without web archives than countries with web archives.

3

Global web archiving endeavors should include multilingual interfaces.

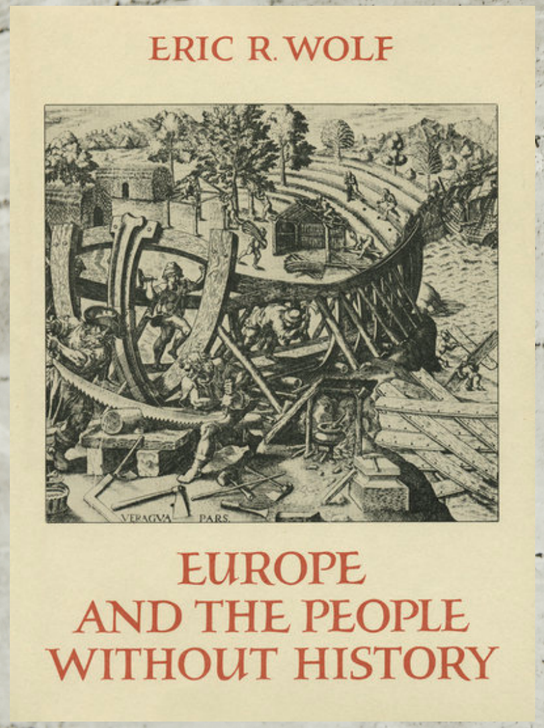
Archives Unleashed + ARCHIVE-IT

VANCOUVER, 22 JUNE 2023

How can we increase the geographic and linguistic diversity of web archiving?

What are the consequences of not doing so?

Risks





LABORATION / IIPC MEMBERS / RESEARCH

The Spanish Web Archive as a training field for Natural Language Processing models

WED, 29 SEPTEMBER 2021 IIPC SENIOR PROGRAM OFFICER 2 COMMENTS

By Alicia Pastrana García and José Carlos Cerdán Medina, National Library of Spain

In the last 20 years most web archives have been building their websites collections. They will be very valuable as the years go by, as much of this information will no longer exist on the Internet. However, do we have to wait that long to see our collections be useful?

The huge amount of information the National Library of Spain (BNE) has built since 2009 has emerged as one of the largest linguistic corpus of current language. For this reason, BNE has collaborated with the Barcelona Super Center (BSC) to create the

Research outcomes

Web Archive Collection, part of Huellas Incómodas Project:
https://idrhku.org/huellasincomodas/webarchive

Spanish Wikipedia entry for web archiving
https://es.wikipedia.org/wiki/Archivado_web

DH Workshop hosted at the Autonomous University of the State of Mexico, 2023
https://idrhku.org/huellasincomodas/investigaciondigital

Online workshops for feminist activists in the State of Mexico: creating an online presence and the importance of web archiving.

Research outcomes

Alan Colin-Arce, Sylvia Fernández-Quintanilla, Rosario Rogel-Salazar, Verónica Benítez-Pérez, Abraham García-Monroy. Web Archiving en español: Barriers to Accessing and Using Web Archives in Latin America, IIPC Web Archiving Conference 2023.

Alan Colin-Arce, Rosario Rogel-Salazar. Creación de archivos web para la investigación digital en humanidades y ciencias sociales. Semana de Humanidades Digitales 2023 https://www.youtube.com/watch?v=oiv0uq-IKAc

$100,000 MXN grant from the Science Council of the State of Mexico under the program Research Funding for Women Scientists.

References

Borgman, C. L. (2015). Big data, little data, no data: Scholarship in the networked world. MIT Press.

Colin-Arce, A., Fernandez, S., Rogel-Salazar, R., Benítez-Pérez, V., & García, A. (2023). Web Archiving en español: Barriers to Accessing and Using Web Archives in Latin America (Version 1). figshare. https://doi.org/10.6084/m9.figshare.22271041.v1

Hargittai, E., & Hsieh, Y. P. (2013). Digital Inequality. In W. H. Dutton (Ed.), The Oxford Handbook of Internet Studies (pp. 129–150). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199589074.013.0007

Millward, G. (2015). I tried to use the Internet to do historical research. It was nearly impossible. Washington Post. https://www.washingtonpost.com/posteverything/wp/2015/02/17/i-tried-to-use-the-internet-to-do-historical-research-it-was-nearly-impossible/

Rogers, R. (2019). Periodizing Web Archiving: Biographical, Event-Based, National and Autobiographical Traditions. In N. Brügger & I. Milligan (Eds.), The SAGE Handbook of Web History (pp. 42–56). SAGE Publications. https://doi.org/10.4135/9781526470546.n4

Wernimont, J. (2013). Whence Feminism? Assessing Feminist Interventions in Digital Literary Archives. Digital Humanities Quarterly, 7(1)



¡Muchas gracias!

@alancolin_arce
@rosariorogel

alan4rce@gmail.com
rosariorogel@gmail.com