

# Lessons from El Paquete, Cuba's Offline Internet

---

*Joel Lenin Pinargote Bravo, Rafael Beto Mpfumo, Luis Alejandro Madruga  
Milanés, Ximena Michelle Cueva, Gretel García Gómez, Amalia Gómez  
Marcheco, Alberto Fernández Oliva, Jeanna Neeffe Matthews, Sam P. Kellogg*

*ACM COMPASS 2018*

*June 20 2018*

# Who Are We/History of Project

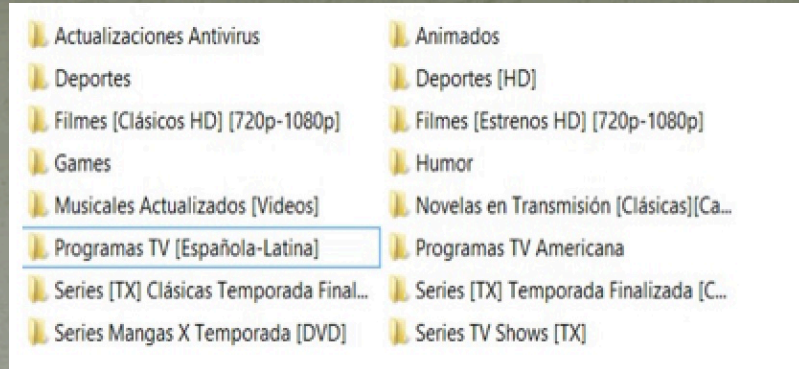
- Began with a class (Universidad de la Habana, November –December 2017)
- After class, added more collaborators (Cuba and US)
- Extended into collaboration with artists! (Cuba and US)



# El Paquete Semanal

- Unique and robust system for distributing digital content on removal media
- The “Weekly Package”
  - Offline Internet alternative
  - Roughly 1 TB of material each week
  - Diverse content from inside and outside Cuba: movies, TV shows, music, sports, applications, magazines, copies of websites including Wikipedia and Revolico, Craigslist-style market place

# Structure



- Realized as a set of 40-60 top-level directories or sections
  - Below top level, max depth we saw was 7; Average 2.76
- Exact list of even top-level directories varies week to week and between distributors
- Editors/curators responsible for acquiring and organizing the content in their section
  - Downloaded by satellite or higher BW access, brought into Cuba on hard drives, produced within Cuba specifically for El Paquete...varies widely with the section and the curator

# Distribution

- Matriz - High-level distributor
  - Assemble section content from editors
  - Distribute on hard drives throughout provinces of Cuba shipped by car/bus/train/plane starting in Havana
  - Substantial role as production studio, adding branding, selling advertizing, sponsoring unique content
- Paqueteros – Intermediate/lower-level distributors
  - Many levels of distribution between matriz and consumers
  - Modify/add/remove content to suit local customers
- Final delivery
  - Copy to consumer's device (hard drive, thumbdrive), copy time substantial
  - ~2 CUC, often freely shared with others

# A few notes

- Not officially sanctioned by government, but clearly tolerated
- “No politics, no pornography” policy
- Distribution system evolved from networks for distributing physical copies of novels and magazines starting in the 1970s and Beta/VHS/DVD movies starting in the 1990s

# Landscape of Alternatives

- WiFi Parks
  - 1 CUC or roughly 1 USD per hour
  - Connection/bandwidth unpredictable; Consumption in public and outside
- Street network
  - Local network disconnected from wider Internet
- Internet access at work/school
- La Mochilla
  - Officially sanctioned “El Paquete”
  - Distributed through Joven Clubs
  - Has foreign entertainment but more of a focus on Cuban content and educational content
- Increasing availability of nauta-hogar/ in-home Internet

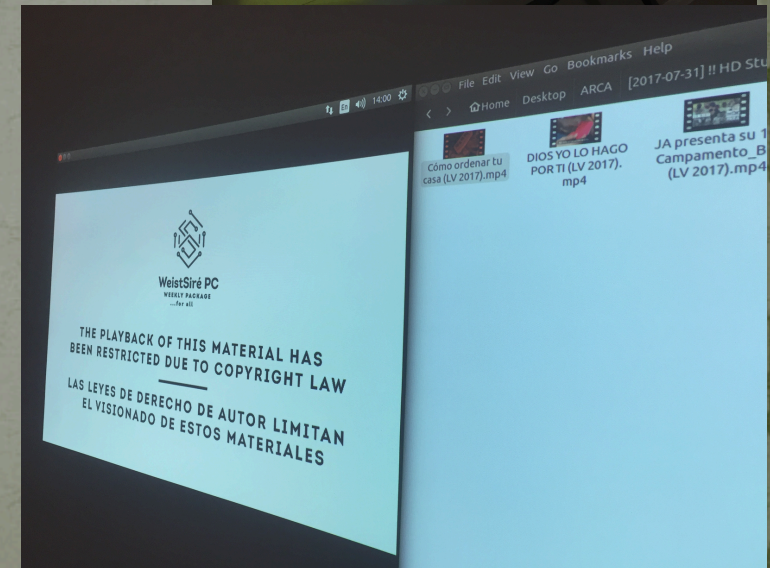
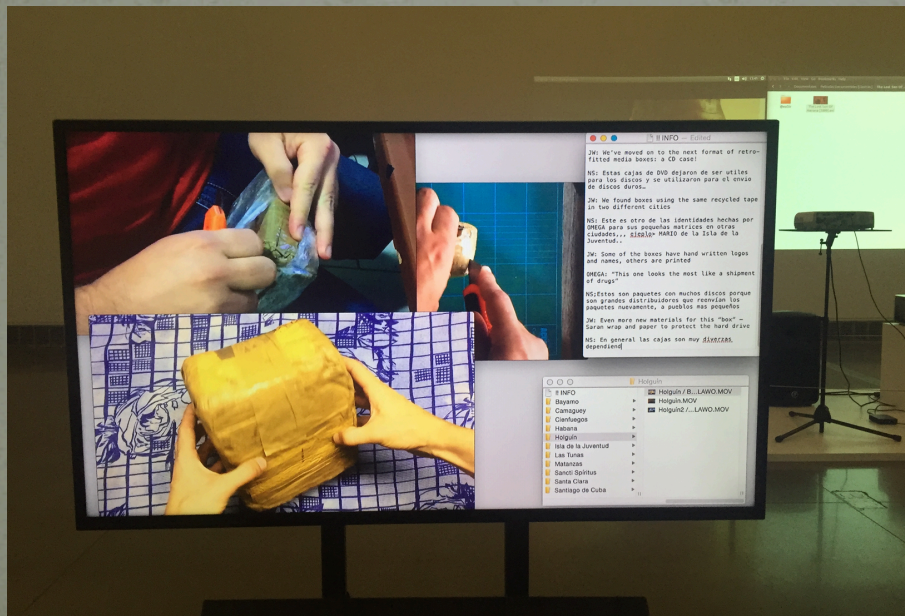
# Our El Paquete Collection

<b>Date</b>	<b>Top-Level Directories</b>	<b>Files</b>	<b>Size (GB)</b>
<b>Dec 4 2017</b>	<b>44</b>	<b>14447</b>	<b>927</b>
<b>Dec 11 2017</b>	<b>47</b>	<b>9196</b>	<b>922</b>
<b>Dec 18 2017</b>	<b>44</b>	<b>10856</b>	<b>931</b>
<b>Jan 1 2018</b>	<b>47</b>	<b>12063</b>	<b>908</b>
<b>Jan 8 2018</b>	<b>46</b>	<b>11512</b>	<b>855</b>
<b>Jan 22 2018</b>	<b>45</b>	<b>10196</b>	<b>924</b>
<b>Jan 29 2018</b>	<b>43</b>	<b>9767</b>	<b>929</b>



# Weist and Siré

- Year long collection
- Acquired from Omega directly
- Queen's Museum



# Content Type

- Video dominates by size; images by number of files
- Variety of video quality/resolutions
- For software, more Android than other platforms

<b>Content Type</b>	<b>Size (GB)</b>	<b>Files</b>
<b>Video</b>	<b>5824 (91.1%)</b>	<b>24540 (31.4%)</b>
<b>Audio</b>	<b>100 (1.6%)</b>	<b>13538 (17.3%)</b>
<b>Images</b>	<b>16 (0.3%)</b>	<b>24757 (31.7%)</b>
<b>Documents</b>	<b>42 (0.7%)</b>	<b>9534 (12.2%)</b>
<b>Android Applications</b>	<b>34 (0.5%)</b>	<b>692 (0.9%)</b>
<b>IOS Applications</b>	<b>11 (0.2%)</b>	<b>128 (0.2%)</b>
<b>Compressed Archive Files</b>	<b>124 (1.9%)</b>	<b>720 (0.9%)</b>
<b>Other</b>	<b>246 (3.8%)</b>	<b>4128 (5.3%)</b>
<b>TOTAL</b>	<b>6396</b>	<b>78037</b>

# Repetition Week to Week

REPEATED CONTENT						
	11-Dec	18-Dec	1-Jan	8-Jan	22-Jan	29-Jan
4-Dec	1274	1823	1332	924	666	376
	36 Gb	22 Gb	10 Gb	1 Gb	3 Gb	1 Gb
11-Dec		1809	1127	875	520	384
		49 Gb	10 Gb	1 Gb	6 Gb	2 Gb
18-Dec			2546	1558	735	505
			17 Gb	7 Gb	6 Gb	2 Gb
1-Jan				2421	1102	535
				40 GB	11 Gb	5 Gb
8-Jan					1485	909
					14 Gb	5 Gb
22-Jan						1241
						13 Gb

- Max week to week repetition was 8.8%
- Between first and seventh sample only 0.1%

# Variations in naming

- Simple predictable variations
  - Addition/subtraction of a quality marker (“Humor” vs “Humor [HD]”)
  - Addition/subtraction of one or more leading “!”
  - Addition of a qualifier for a genre of content (e.g. “Películas por Género [HD] [Fantástico]”)
  - Indication of specific actor (e.g. “Películas por Actores {Gwyneth Paltrow}) or year (e.g. “Trailers [2016]”)
  - Plural v.s. singular (e.g. “Deporte” vs. “Deportes”)
  - With and without accent marks (e.g. Música vs. Musica)
- Intuitive changes but indication of manual modification/lack of programmatic structure
  - Inconsistent use of grouping symbols [], {}, ()

# Lessons

- Opportunity to reconsider our mix of online/offline consumption
  - Robust distribution networks based on human relationships
  - Resilient to disruptions in electrical power, end-to-end connectivity and other infrastructure
  - Offline by choice even when connection available
  - More anonymous? Less trackable? Differently trackable
- Distribution of educational/research materials
  - Consider modifications to MOOC platforms

# Future Work

- Collaborations with Weist and Siré
  - Detailed analysis of year long collection
  - Categorizing content by country of origin, date of original production, time from debut to appearance in El Paquete
  - Parsing names to extract titles, actors/musicians, years
- Study of material added/removed/changed along distribution system
- Long-term speed/accessibility measurements