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 Neefe 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
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 advertising, 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

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

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

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

- 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. “Peliculas por Genero [HD] [Fantastico]”)
  - Indication of specific actor (e.g. “Peliculas 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