Walking the path between archival activism and the social archive(s)

Vicenç Ruiz

Society of Catalan Archivists and Records Managers
1) Infosphere and surveillance capitalism
2) Archiving Activism
3) The path to the social archive(s)
   - From data to information
   - From the social web to the social evidence
INFOSPHERE AND SURVEILLANCE CAPITALISM
“All societies (including the archivists residing in them) assign greater or lesser value to different dimensions of the three-way interplay of social structures, societal functions, and citizens and groups [...]. Macroappraisal theory posits that archivists may determine such societal values by specifying the generic functional attributes and points of special intersection, conflict, or interplay among the creators of records (that is, structures, agencies, actors); sociohistorical trends and patterns (that is, revealed through functions, programs, activities); and clients, customers, citizens, or groups on whom both function and structure impinge and who in turn influence both, directly or indirectly, explicitly or implicitly”

Terry Cook (2006)
Which sociohistorical patterns are revealed by the creation of social media records?

<table>
<thead>
<tr>
<th>Network Society</th>
<th>“a society whose social structure is made up of networks powered by micro-electronics-based information and communications technologies” (Manuel Castells)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infosphere</td>
<td>“the whole informational environment constituted by all informational entities, their properties, interactions, processes, and mutual relations” (Luciano Floridi)</td>
</tr>
<tr>
<td>Surveillance capitalism</td>
<td>“new form of information capitalism that aims to predict and modify human behaviour as a means to produce revenue and market control” (Shoshana Zuboff)</td>
</tr>
</tbody>
</table>
Archives are no longer the by-product of the activities of organisations but a means of production, accumulation and control:

“The extractive processes that make big data possible typically occur in the absence of dialogue or consent […] Populations are the sources from which data extraction proceeds and the ultimate targets of the utilities such data produce” (Zuboff, 2015)
So, what to do as archivists?

• The monopoly of archiving held by states, large corporations and social elites is only apparently over

• Democratizing archival knowledge is essential to turn big data into social counterpower

• The social (media) archive(s) are not about democratically building our future social memory, but socially building our present democracy
ARCHIVING ACTIVISM
A definition: “an archivist or archival institution, whether formal or independent, acting to collect and document political, social movement and other activist groups and campaigns” (Andrew Flinn & Ben Alexander, 2015)

An example: DocNow

<table>
<thead>
<tr>
<th>DocNow</th>
<th>Tweet Catalog</th>
<th>Hydrator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appraise Twitter for archival and research collections.</td>
<td>A catalog of publicly shared tweet ID sets. Add yours!</td>
<td>&quot;Rehydrate&quot; tweet ID sets into tweets with metadata.</td>
</tr>
<tr>
<td>Social Humans Labels</td>
<td></td>
<td>Diff Engine</td>
</tr>
<tr>
<td>Labels for ethically describing and sharing social media data.</td>
<td>Twarc</td>
<td>Track changes in news articles through RSS feeds.</td>
</tr>
<tr>
<td></td>
<td>Archive Twitter JSON using this command line tool.</td>
<td></td>
</tr>
</tbody>
</table>
# NoTincPor
Creator: Society of Catalan Archivists
Tweets: 52,808
Added: 2017-09-26 12:30:45 +0000
Published: 2017-09-26
Date Coverage: 2017-08-17 - 2017-08-26
Tags: barcelona, cambrils, attack, terrorism

**Description**

2017 Catalonia attacks were a terrorism action against pedestrians who were at La Rambla (Barcelona) and beach promenade of Cambrils on the afternoon and night of 17-18th August 2017. We selected #NoTincPor hashtag because it was the motto of the demonstrations during that period and was the most positive message. No one knew how the situation would go so the best option was to collect the dataset using the search API and limit by time between August 17 - 26, which was the final demonstration in Barcelona for the victims.

**Catalan Referendum**
Creator: Vicenç Ruiz Gómez, Aniol Maria, Ed Summers
Tweets: 8,410,431
Added: 2017-10-17 18:25:22 +0000
Published: 2017-10-17
Date Coverage: 2017-09-19 - 2017-10-5
Tags: Spain, Catalonia, activism, politics

**Description**

This dataset contains identifiers for 8,410,431 tweets that were collected between September 19, 2017 and October 5, 2017 that mentioned CatalanReferendum, #CatalonianReferendum, #Catalonia, #1oct, #1o or #votarem. These hashtags were used in the lead up to the Catalan Independence Referendum on October 1, 2017. The referendum was declared illegal under Spanish law, and the Spanish police attempted to prevent it. The data collection was a collaboration with Vicenç Ruiz Gómez and Aniol Maria of the Society of Catalan Archivists working in conjunction with Ed Summers of the Maryland Institute for Technology in the Humanities. The hashtags were selected after monitoring the CatalanReferendum hashtag for several hours on September 28 to determine what the top hashtags being used were. The tweets themselves were collected from the Twitter Search API using twarc and its twarc-archive utility. twarc-archive was run every hour to collect the tweets that occurred since the last run.
Documenting The Now Recommendations:

• Archivists should engage and work with the communities they wish to document
• Documentation efforts must go beyond what can be collected without permission from the web and social media
• Archivists should follow social media platforms’ terms of service where they are congruent with the values of the communities they are attempting to document.
• When possible, archivists should apply traditional archival practices such as appraisal, collection development, and donor relations to social media and web materials

Ethical Considerations for Archiving Social Media Content Generated by Contemporary Social Movements: Challenges, Opportunities, and Recommendations (April 2018)
Understanding *Social fonds*

- **Definition:** “the entire body of records created by a community of users through a participatory dynamic in social platforms around a shared interest or event”

- **Characteristics (vs traditional fonds):**
  - Produced by a collective creator (their only bond may sometimes be as platform users)
  - Produced in a privative system not owned by creators (digital colonialism)
  - Produced massively and briefly

- **Priorities (according to those characteristics):**
  - To design a participatory recordkeeping model > Community archives, Monash PRM
  - To act proactively so as to guarantee a democratic and free long-term access > Activist data archiving
  - To develop tools that allow to automatize the analysis (appraisal, description,...) and exploitation of information > Computational Archival Science
THE PATH TO THE SOCIAL ARCHIVE(S)
The origins of #Cuentalo:

• 26 Apr 2018 (Sentence and a huge social response)
• 27 Apr 2018 (A newspaper article and a tweet)

“I have started the day proposing #cuéntalo because I think that ALMOST ALL OF US have suffered some kind of sexual aggression. THREAD”
THE PATH TO THE SOCIAL ARCHIVE(S)

FROM DATA TO INFORMATION
“... *data* are some combination of elements of raw content, such as numbers or letters, and *information* is contextualized data, or data infused with layers of meaning. A *record* captures information or data in a fixed medium; it is a “whole” thing: an email, a report, or a text message. *Evidence* is any source of information that provides demonstrable proof. We cannot say that a Facebook post is only information, or a photograph is always evidence, or a data base is just data. If the source – data element, photograph album, or membership database– can be used to provide proof of actions, transactions, or decisions, then it has evidential value”

Laura Millar (*A Matter of Facts*)
Data archiving:

Aniol Maria started the capture on April 30, finishing on May 16 (when the HT became marginal):


• twarc filter "’#Cuéntalo",#NoEsAbusoEsViolación > Cuentalo-filter-2018.04.27-2018.05.13.jsonl

Result: 2’1 million tweets
Building a Community of Practice

Cristina Fallarás (journalist)
Fernando Cucchietti (BSC-CNS)
Karma Peiró (data journalist)
Aniol Maria (AAC)
Vicenç Ruiz (AAC)
The BSC-CNS Scientific Visualization Team

Data pre-processing
- Dynamics of the HT; Geolocalization of tweets; Cleaning tweets texts; Vectorization of words

Generating a training dataset
- Volunteers (“us”) read/tagged 10 thousand tweets
- Questions: who (1st person, 3rd person, Support, Against, Others); what (murder, rape, sexual assault, abuse, harassment, fear, anger/anguish)

Training Neural Networks
- Word Embedding (GloVe: unsupervised learning algorithm for obtaining vector representations for words)
- Cross Validation
- Neural Network Architecture

For a deeper explanation: [https://www.youtube.com/watch?v=0WM2wlvY-LA](https://www.youtube.com/watch?v=0WM2wlvY-LA) (AAC, XVII Conference, May 2019)
#Cuentalo: some figures
The design of an interactive data visualization: *A constellation of Horror*
When dataviz shows us also the recordkeeping design...

Fig. 3 A Participatory Recordkeeping Continuum Model
THE PATH TO THE SOCIAL ARCHIVE(S)

FROM THE SOCIAL WEB TO THE SOCIAL EVIDENCE
NEXT STEPS (funding needed)
1. DEVELOPMENT AND IMPLEMENTATION

- GOALS:
  - Testing the participatory recordkeeping model for the social web
  - Facing the challenge of the preservation and access to the “social fonds”
  - Providing a tool for analysing and exploiting massive data

- DELIVERABLES:
  - An initial “archive’s community” -formed by creators, archivists, engineers, ...- around geolocated tweets of a city/area (for example, Barcelona)
  - A trusted repository for datasets and social web content (Dataverse, B2safe,...)
  - A participatory recordkeeping platform (Mukurtu, ...)
  - A defined digital ethical framework for archiving the social web
  - An open source tool for analysing massive data (BSC-CNS)
  - Technical documentation (guides, toolkits,..)
NEXT STEPS (funding needed)

2. INTEGRATION AND PLURALIZATION

• GOALS:
  – Providing institutional sustainability and political mandate for the archiving of social funds
  – Promoting interdisciplinary research
  – Pluralizing the theoretical and technological knowledge generated by the project
  – Guaranteeing archival autonomy of citizens as a tool for political empowerment

• DELIVERABLES:
  – Integration of the participatory platform into an institutional archival system, in order to start a regulated policy of capturing datasets and contents of the social web
  – Federation of nodes of other geographical areas (Catalonia, Spain, Latin America) based on the experience of the initial archive’s community
  – Interdisciplinary research projects in academic and professional fields
  – Workshops and seminars (hackatons, conferences, ...)
  – Preparation of pedagogical and outreach materials
REMEMBER:

ARCHIVAL RAGE AGAINST THE MACHINE
#StandUpForCatalonia

#NoEsJusticiaEsVenganza

MOLTES GRÀCIES!
THANK YOU!

www.arxivers.com