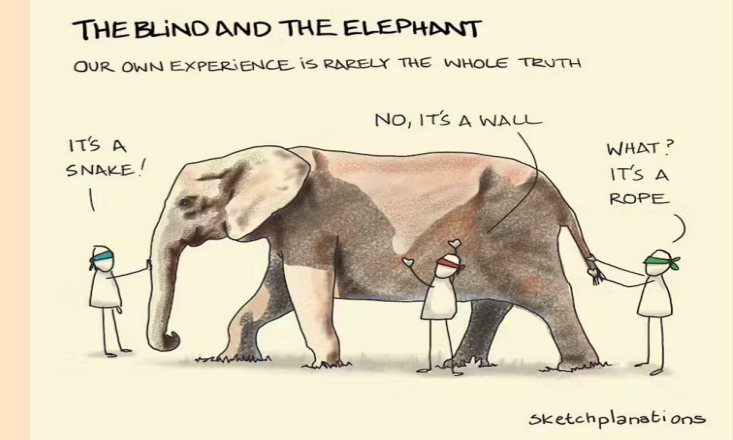




2024 Paris Olympic games: a lens to analyze YT algorithmic recommendations

- a one off-event, likely to dominate the media spotlight and narrative,
- laden with political and sport significance
- often presented as a "unifying event"
- => do YT algorithms convey a shared narrative or distinct narrative and values?
- => can we build an archive reflecting diverse perspectives on the same subject?

Why ?



Yvette Assilaméhou-Kunz, Université Sorbonne Nouvelle
 Michaël Attali, Université Rennes 2
 Dorothée Benhamou-Suesser, National Library of France
 Amélia Ferreira, BnF et LAAS-CNRS
 Erwan Le Merrier, Inria, Rennes
 Julien Mésangeau, Université de Lille
 Gilles Tredan, LAAS-CNRS

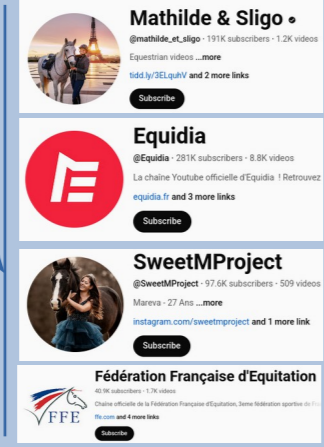
... through an exploratory approach: quali/quant, diverse disciplines:
 National Library of France and researchers in Computer Sciences,
 Sport History, Communication and Information Sciences, Sociology

Web Archiving practices' blind spot
 - Algorithmic recommendations are shaping our access to information but remain absent from institutional Web Archives:
 - can we capture both the personalization of the browsing interfaces and experiences offered by RSM platforms and the personalization of narratives and perspectives induced by them?

How ?

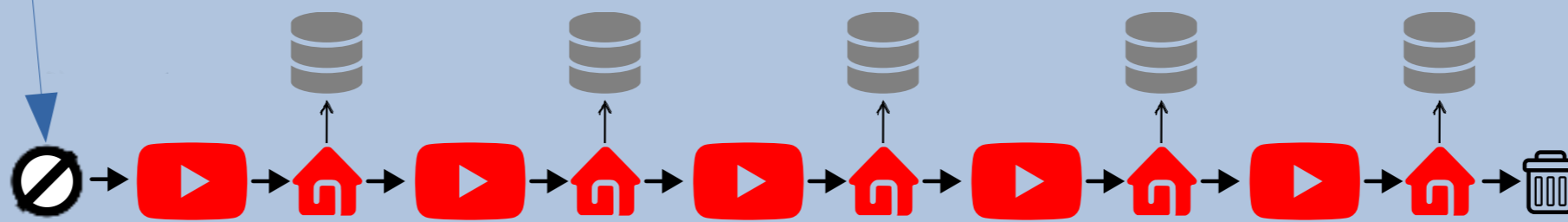
Profiles HOW ?

- Predetermined profiles, "Reasoned arbitrariness"
- Spanning the spectrum + polar pairs
 - Media profiles
 - Mainstream/Indépendant, left-wing / right-wing
 - Sports profiles:
 - General: elite/popular
 - Specific: Fitness/Horseback riding, basketball
 - Miscellaneous and political
 - masculinism vs Féminism
 - Transphobic, Ecologiste
 - Top10 as reference profile, Parisien, Provincial



Subjective Corpus: "SockPuppet" Audit

Bots emulating 16 user archetypes profiles
 = Video consumption profiles, bots with diverse browsing history
 Bots follow a personalized web browsing path (viewing first videos from YT channels associated with their profile)



Collection Campaign

Dates : 07.09 - 11.08.2024

"Objective" Corpus:

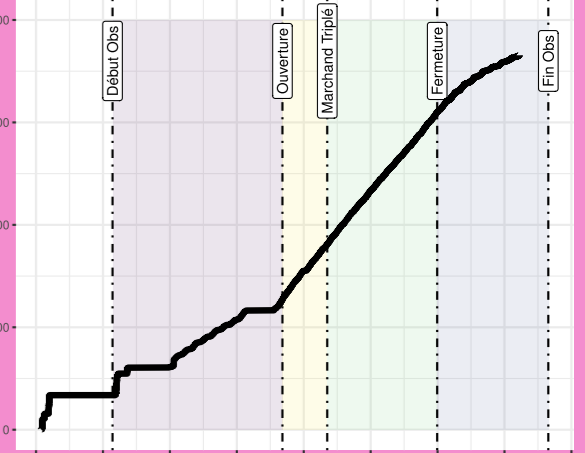
- (nearly) all videos matching "jeux olympiques" from the last hour, fetched every 30 minutes.
- All matched videos are then monitored daily

Archiving and Analyzing YouTube Recommendations during the Paris 2024 Olympic Games

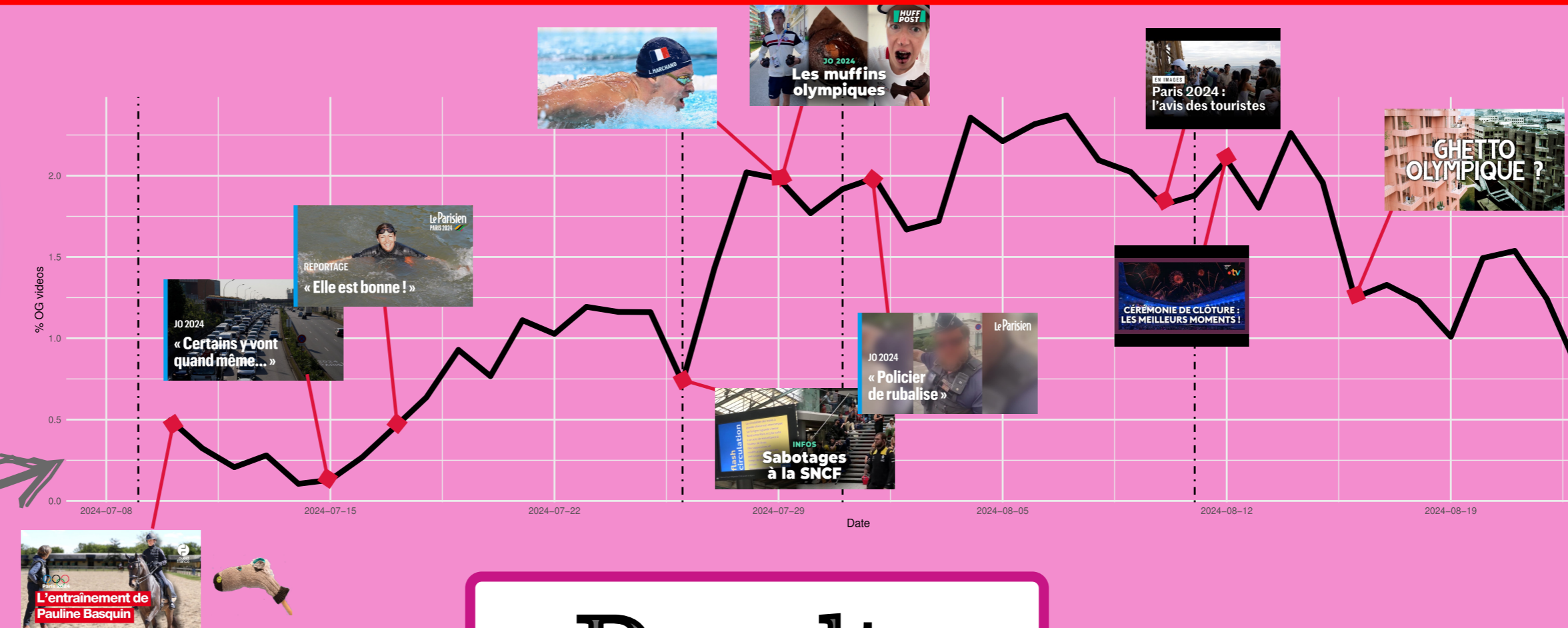
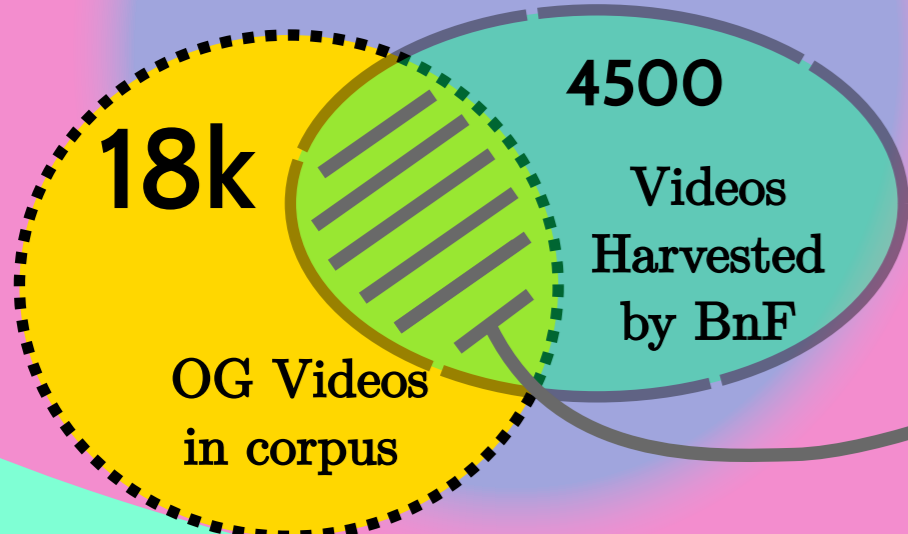
Profile recommendation similarity as a map:



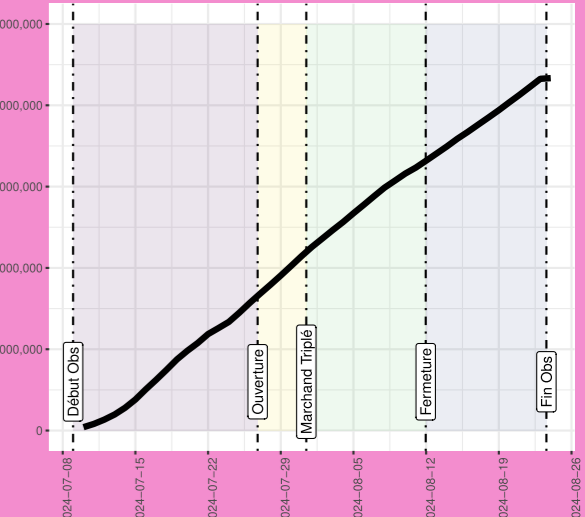
OG 'Objective' Corpus size



21.7m
 All collected recommendations



Number of collected Recommendations



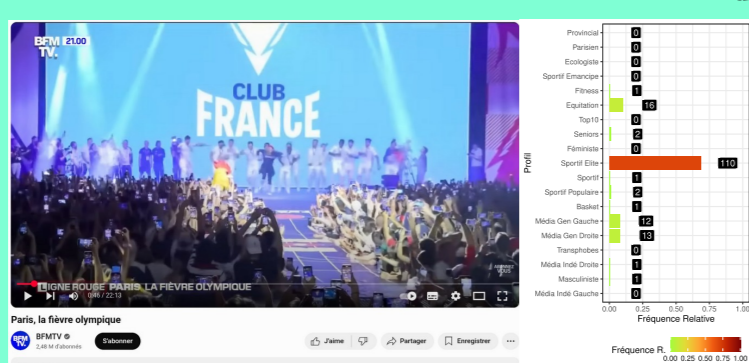
Results

Personalization Representativity ?

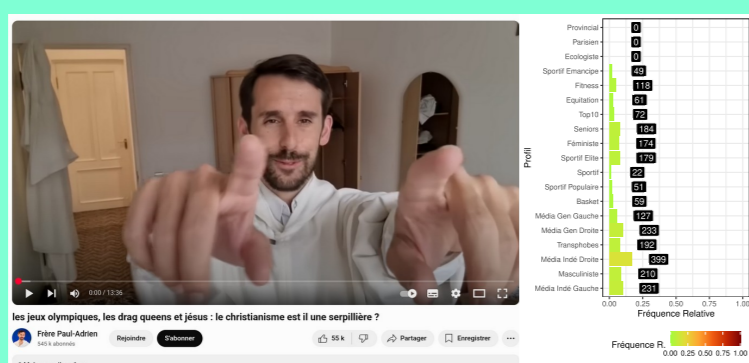
"Specific" universalisms*

Most common recommendations are hand labelled with their values

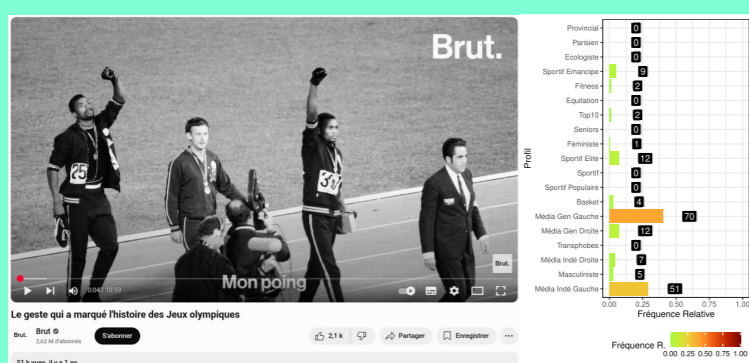
Universalism as Unity/Beautiful World



Universalism as Unity (inter/intra national) Peace, Wisdom



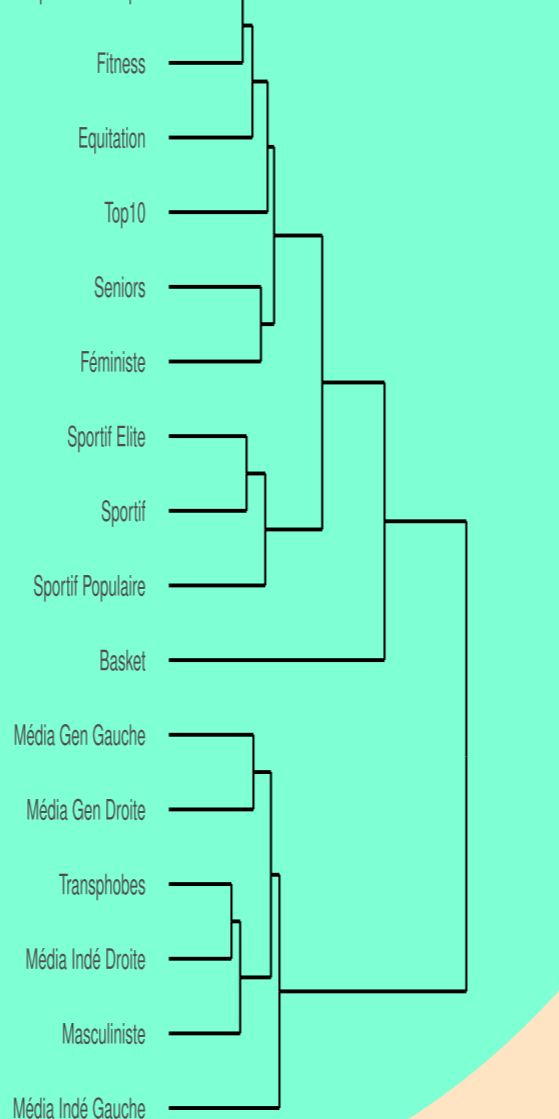
Universalism as Peace, Social Justice



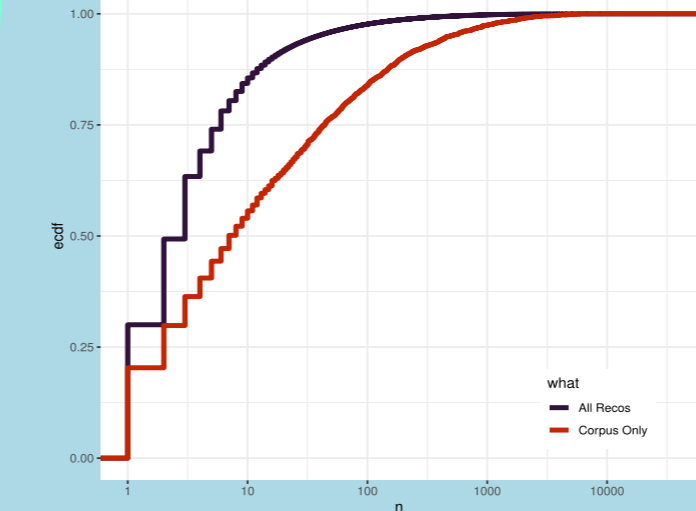
Profiles can be described using values distribution

.. that translates into intra-profiles values similarity

.. that translates into intra-profiles values similarity



Fraction of the 268/18k (15%) corpus videos recommended at least once

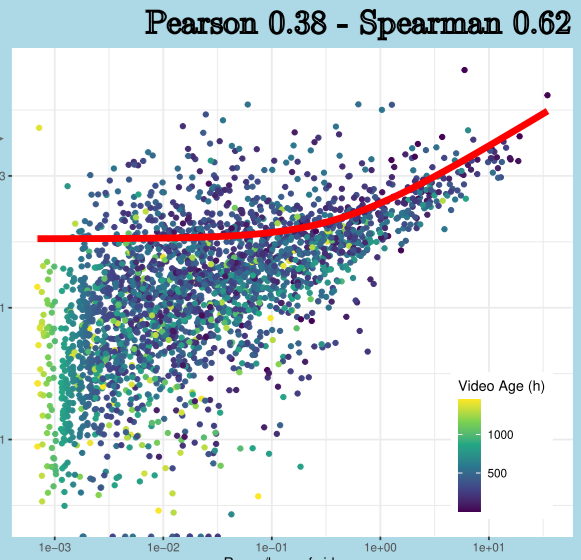


Heavy tailed recommendation distribution: 85% of corpus OG videos are never recommended

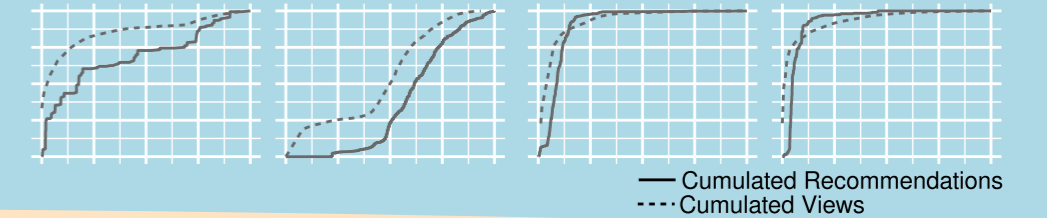
Assuming a "Neutral" Recommender, Reco/Views correlation translates into personalization coverage measure: "Roughly" we cover 62% of personalization space

This perspective is wrong in many ways But it is the first personalization coverage measure we know about

Recommendation/Views Correlation Pearson 0.38 - Spearman 0.62



Views and Recommendations: A non trivial relation



Conclusion

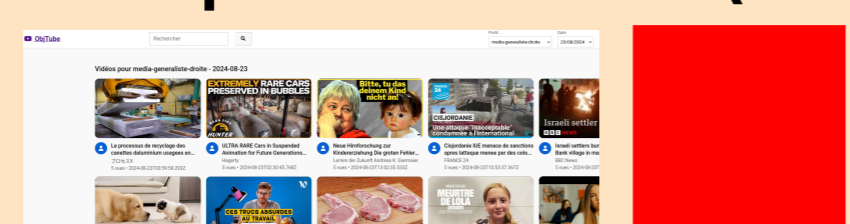
Questioning web archiving practices?

- A metric driven approach for certain curated crawls (most viewed, most recommended) complementing the usual selection criteria based on a typology of actors (eg. teams, athletes, officials, NGO, opposant, etc.)
- Complementing "content producers" approach with a "consumer centered" approach?
- Replaying the harvested recommended video: a guided tour in BnF Wayback

Research questions

- YT algorithm do provide different narratives on the same event
- media content dominates OG top recommended videos, OG videos are rarely sport-related -- rather "outside the stadium"
- Different values are exposed across profiles; Identical values may refer to different notions (eg Universal)
- Some surprises: eg. Feminist profile fail, why?
- Double vertigo: avalanche of metadata -- no apriori meaning.
- Abundance of content, yet little information, mostly of anecdotal nature
- Our analysis is very restricted: this large dataset will hopefully serve exploring many other issues

Live Exploration interface ! (beta)



*: Cf Schwartz, Shalom. "A theory of cultural value orientations: Explication and applications." Comparative sociology 5.2-3 (2006): 137-182.

