The paper proposed for Cultural Data Analytics Conference 2023 is based on the research project Crossing Borders Archives[1] (CROBORA), funded by the French National Research Agency, 2021-2024, led by the Sic.Lab at Université Côte d'Azur. The general objective of the project is the collection and analysis of the reuses of broadcast archives within television programs and web videos. The project deals with images linked to the construction of the European Union. In this proposal, our objective extends beyond merely describing and analyzing the data collected by the CROBORA team, but also an interactive visualization platform, enabling users to explore the data in accordance with their specific requirements.

The data cover television in France and Italy, including the French Legal Deposit of and Television (*Institut national de l'audiovisuel/INA*), the archives of the Italian public service of television (*Radiotelevisione italiana/Rai*). The focus is on news broadcasting from 2001 to 2021 from major national channels, including TF1, France 2, and ARTE from France and Rai 1 and Rai 2 from Italy. For the web data collection, two approaches were used: INA's legal deposit of the archived web and straight from the live web, using a textual search in the INA database, search engines of the live web, as well as an artificial intelligence software *Snoop* which specializes in automated visual detection. Selected accounts included heritage institutions, EU official accounts and media accounts from YouTube, Dailymotion, Twitter, and Facebook, with a focus on their published content from 2001 to 2021. The keywords used for search queries were "*archiv*/repertorio*" and "*europ**", aiming to gather most stock images concerning the European Union. The total occurrences of stock shots in our corpus is 27,913.

All data is conveniently available on the CROBORA Visual Platform[2], a user-centric tool that features integrated, robust functions that facilitate in-depth data analysis, enhancing the overall user experience and utility. The platform allows users the freedom to traverse through the corpus by amalgamating different keywords. Additionally, it offers the capability to conduct exploratory data analysis utilizing two distinct functions: 'Distributions' and 'ARViz'. We have incorporated a software called Muvin[3] within the 'Distributions' function of our platform. Muvin is an innovative visualization technique that blends streamgraphs and node-link diagrams to facilitate the navigation of multidimensional collaboration networks. Through Muvin, users can delve into the co-occurrence of keywords within the CROBORA data and trace the chronological progression of their usage. As for the ARViz[4] function, it's designed to address the issue in association rule mining in CROBORA corpus. Our user interface champions data exploration, providing a comprehensive overview of rules via a scatter plot, a detailed view of rule subsets through a chord diagram chart, and itemsets via an association graph.

In this paper, we hypothesize that the digitization of audiovisual data has precipitated a trend towards visual homogenization in the media space, most notably in the news broadcasting sector. We aim to undertake several key objectives: Firstly, we plan to describe the challenges we've encountered in the process of collecting cultural data. In particular, we will discuss the stark differences between web data and television data, specifically regarding their metadata and the volume that researchers need to manage. Secondly, we will demonstrate certain recurring patterns in both metadata and visual data themselves through pertinent examples of data visualization using various quantitative tools, such as R, Gephi, and Iramuteq. Lastly, we

---

[1] https://crobora.hypotheses.org/
[2] http://dataviz.i3s.unice.fr/crobora
[3] http://dataviz.i3s.unice.fr/muvin/
[4] http://covid19.i3s.unice.fr:8080/arviz/

will introduce the data visualization platform of CROBORA. By doing so, we hope to shed light on potential future research opportunities for the community interested in audiovisual archives concerning the European Union.
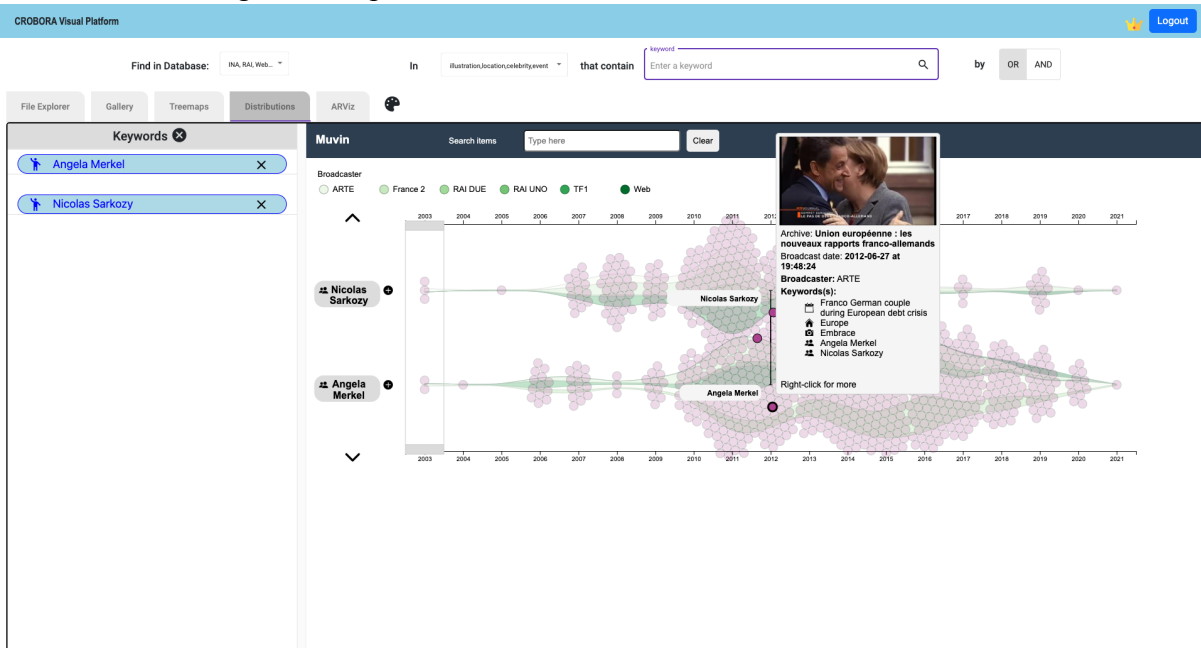

Illustration 1. Combination of keywords "Angela Merkel" and "Nicolas Sarkozy" using 'Distributions' function
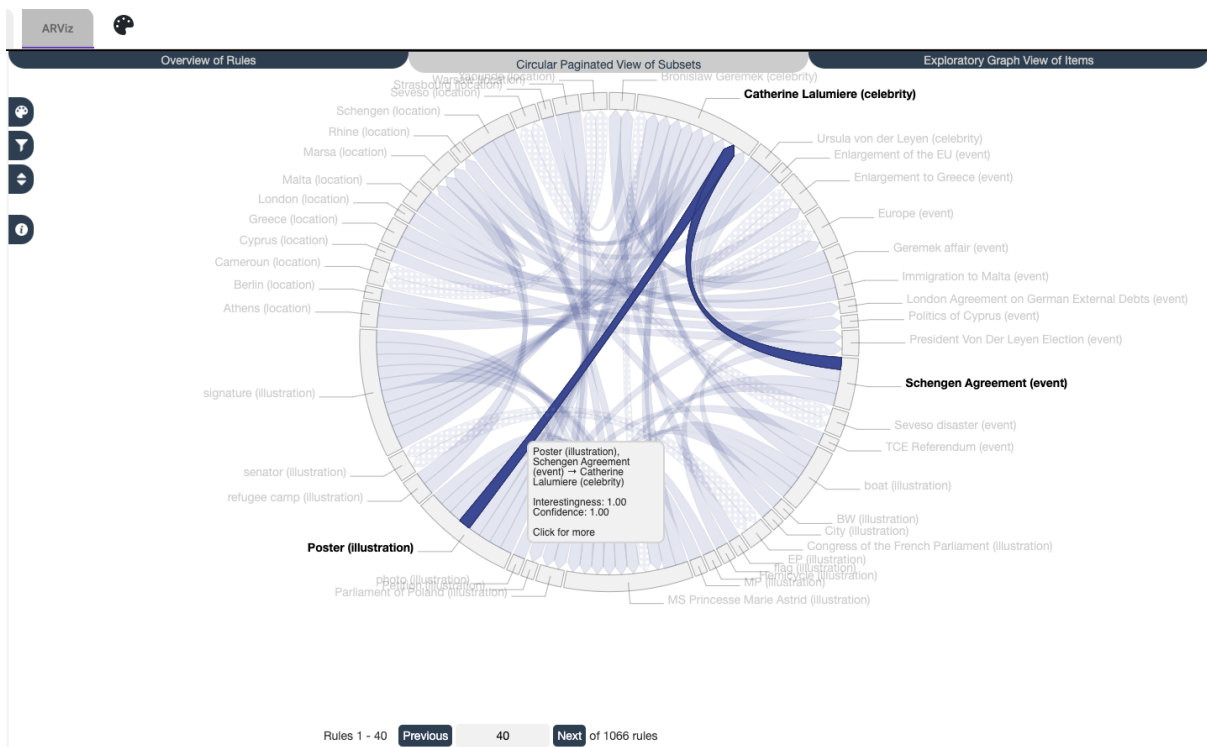

Illustration 2. Some association rules in CROBORA corpus using 'ARViz' function

Please note, this paper is an expanded version of prior research, of which certain results have been recently accepted in two peer-reviewed journals. The two articles should be published in September 2023.

Shiming Shen, Matteo Treleani, Dario Compagno, Marco Winckler, 2023. "From Stock Shots to Ghost Data: Tracking Audiovisual Archives about the European Union". *VIEW Journal of European Television History and Culture*, in press.

Matteo Treleani, Shiming Shen, Aline Menin, Marco Winckler, 2023. "Circulation et répétition des images de stock dans les médias audiovisuels. La représentation visuelle des migrants dans les JT européens (Circulation and repetition of stock images in audiovisual media. The visual representation of migrants in European TV news)". *Les Cahiers du Numérique*, in press.