

The request:

The Authoritarian Literature Archive (ALA) is asked by the Centre for Linguistic Analysis (CLA) if they would be interested in collaborating on the **H2020 Project** “EVO-NALA: EVOLutive NATural Language Assessment” which focuses on **researching common language patterns in nationalist and fascist European literature** from the 1920’s and identifying similar language patterns in internet posts from the 2020’s.

Digitization:

If the Authoritarian Literature Archive (ALA) decides to collaborate on this project, **they would be required to digitize Collections 123-A and 123-B** (Europe, Political Science, National Syndicalism, 1919-1929). However, the ALA is a publicly funded archive and **does not have the necessary equipment**, so they would have to **rely on the infrastructure of the CLA to perform mass digitization of the two collections**.

Data mining:

Neither of the two organizations possesses the means to perform **automated text processing and statistical analysis (text and data mining)**, so, the digitized books would have to be handled by an outsourced partner, a computer science organization known as the Boolean Logic Association (BLA).

The data scientists at the BLA will manually tag the digitized books as well as twelve thousand right-wing sub-forums posts from the website Saiditt. Then they will run the annotated text through a neural network to find correlations.

Copyright concerns:

The ALA asks you about the lawfulness of this operation since this is not part of their common activity. They are concerned because the two collections that the CLA needs to digitize are **mostly composed of in-copyright works**. They want to know if they can digitize all the books from that collection inside the premises of the CLA, and if they can **authorize the data scientists at the BLA to perform reproductions and retain copies of those works too**.

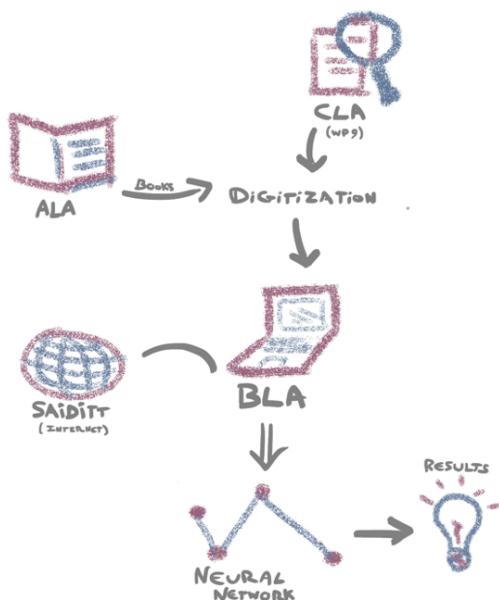
Contractual and technical restrictions:

The ALA also noticed that the terms and conditions on Saiditt’s website mentioned that “It is prohibited to access, search, or collect data by any means (automated or otherwise)”. They suspect that, since **automated web-scraping is restricted by Saiditt’s terms and by machine-readable means** (robots.txt), automatically downloading the content of several sub-forums may be unlawful and they don’t want to be “involved in any shady stuff”.

What would your response to ALA’s inquiries about digitization, text mining, the copyright of the works and lawfulness of web-scraping be?

For the purposes of this exercise, assume the country of all the legal persons has implemented Articles 3 and 4 of the DSM Directive “as they are”. You can use the suggested materials as well as the internet to complete this task.

Please submit your answer in the Open Observatory (link provided in the room), in the form of a Proposal, titled: “## - The Authoritarian Literature Archive”, with “##” being the Training ID Number (A1, A2, A3, etc...) that was attributed to you (card).



The request:

The Local Archival Museum Unit (LAMU) is preparing a new project that outlines and narrates the history of your city, as a commemoration of the 100th anniversary of the Liberation Revolution, a movement of insurgence against authoritarian control in the country. For this purpose, they are collecting various types of works from two different periods: 1901-1921, the twenty years of oppression that precede the revolution’s eruption; and 1922-1952, the thirty years of enormous technological and cultural progress that followed the liberation of the people.

Materials and collection methods:

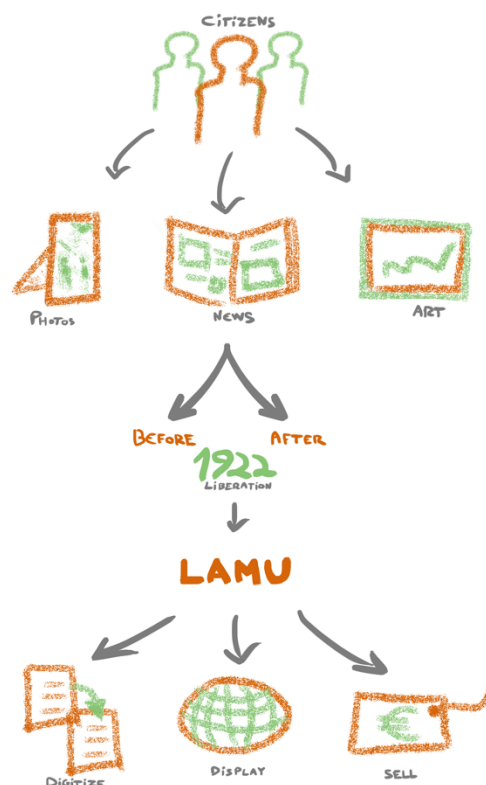
For this digitization and preservation task, they request citizens to search for items to digitize that can help tell the untold history of the Liberation Revolution to the world. Anything that can help the curators map out daily life before and after the Revolution is accepted, but priority is given to photographs, newspapers, paintings, writings (inc. correspondence), and propaganda posters. The collection methods that LAMU uses to obtain works from citizens are as follows: private donations (permanent); deposit and temporary loan (for digitization and temporary exhibition), digital surrogate (asking the citizen to send pictures of the artifacts, objects or works).

Exhibition and merchandising:

For the Centenary, LAMU will have a free exhibition, and will publish low-res versions of all the works in their open website. Citizens also have the chance to vote on five iconic submissions, which LAMU will mint as NFTs, available on online auction. The profits will be used to cover the running costs of the Centenary.

Copyright concerns:

The LAMU approaches you for your advice on the matter. They want to make sure that they can digitize and preserve the historical works they have in their possession and display them publicly and online. Regarding the copyright status of the donations, they are mainly concerned about three aspects: all works from the second period are in-copyright; some of these in-copyright works have no mention of the author, which could not be discovered; some propaganda posters from the second period were never meant to be commercially available. Finally, they are also not sure about capitalizing on citizen submissions to mint NFT’s, even though the profits will be used to finance the preservation efforts too.



What would your response to LAMU’s inquiries about digitization, preservation, orphan works, out-of-commerce works, and public domain works be?

For the purposes of this exercise, assume the country of all the legal persons has implemented the DSM Directive “as is”. You can use the suggested materials as well as the internet to complete this task.

Please submit your answer in the Open Observatory (link provided in the room), in the form of a Proposal, titled: “## - The Revolution Centenary”, with “##” being the Training ID Number (B1, B2, B3, etc..) that was attributed to you (card).