Measuring the Impact of Digital Culture

inDICEs

Deliverable 1.5

Report on data gathering v3

inDICEs

Measuring the Impact of Digital Culture

D1.5 – Report on data gathering v3

Version 1.0

Authors: **Maria Tartari, FBK-IULM**

    Federico Pilati, FBK-IULM

    Katinka Böhm, WLT

    Elisa Pellegrini, EFHA

    Marco Rendina, EFHA

    Marta Franceschini, EFHA

    Corinne Szteinsznaider, Michael Culture

    Marzia Piccininno, Michael Culture

Contributing Partners: **Pier Luigi Sacco, FBK**

    Oriol Artime, FBK

    Riccardo Gallotti, FBK

Reviewers: **Franco Niccolucci, PIN**

    Marzia Piccininno, Michael Culture

Document History

- 22.11.2021 – Document Created from Template

- 22.11.2021 - Index definition

- 01.12.2021 - Contribution on behavioural analyses

- 14.12.2021 - Internal review

- 18.03.2022 - Partners review

- 24.03.2022 - Updating

- 29.03.2022 - Last version

Table of Contents

1. Executive summary

This deliverable describes the data collection processes and status, and the datasets gathered in the months M21-24 by WP1. In this period, the data analysis focused on the second part of the lists of case studies collected by the inDICEs partners about the CHIs' social networks users' behaviour, especially within the Covid-19 wave of forced digitization. Moreover, the behavioural analysis regarding Wikipedia and TikTok users and about digital content co-creation has been finalised and summarised in the present deliverable, with the aim of enriching one of the most relevant inDICEs theoretical frameworks, namely the theory of the 3.0 Culture, which regards the contemporary strategies and dynamics of cultural production and reproduction in the web 2.0.

The inDICEs data collection processed and/or stored consists of:

- building and analysing a set of case studies, composed by lists of Facebook and Instagram accounts, per macro-sectoral areas of cultural and creative institutions. The sectors' case studies that have been identified and analysed are:
    - European Fashion GLAMs;
    - European National Theatres;
    - European Archeological Sites.
- early results on users behaviour in content co-creation of Wikipedia platform and TikTok social network;
- repositories of metadata-enriched documents gathered from web sources and through social media APIs depicting the public news and online debate with a focus on content published by cultural heritage institutions and in the cultural and creative sector;
- a Knowledge Graph used to store factual and semantic knowledge on entities specific to the inDICEs uses cases, e.g. GLAM institutes.

This data was gathered with the purpose to:

- detect trends regarding the levels of digital cultural participation of the most used social network sites, with a focus on the impact of Covid-19 wave of forced digitization on users' behaviour and CHIs behaviour;

- enrich inDICEs repositories with sources gathered from the Web and from social media;

- supporting the definition of the inDICEs Policy Guidelines due by the end of the project.

2. Objectives

The objective of this deliverable, which is the third inDICEs data gathering periodical report, is to describe the status of data gathering activities, with specific information on the methodology, quality, reliability and accessibility of the information gathered during months 21-24 by WP1.

The report is aimed at describing the results achieved by this work package in this specific time window (2019-2021), as well as to give an overview of the preliminary research conducted. Namely the analyses carried out are based on a set of case studies differentiated per Cultural and Creative Sector (CCS). This selection was part of a broader consultation with field experts and inDICEs partners.

The aim of the present data collection is to proceed with an analysis of the European sectors of the National Theatres, Fashion GLAMs and Archeological Sites. The main goal is to detect the trends regarding the levels of digital cultural participation of the users of the different sectors' organisations. This analysis has been conducted with a specific focus on a temporal window that can help make a comparison of the relationship between users and CHIs' digital platforms before, during and after the pandemic.

This is the third of four periodic reports about data gathering. It outlines the status of the first phase of the inDICEs project, namely the data gathering activity. The data gathering strategy of M 21-24 of work has been devised to collect new data on CHIs digital platforms' users behaviour and, on the other side, to enrich inDICEs repositories with sources gathered from the Web and from social media. In particular, strategies of acquisition of relevant data through Facebook and Instagram have been defined in a previous deliverable (D1.3) and implemented during the first trimester of 2021.

In order to express how this data collection and analysis support the inDICEs objectives and activities, it is important to underline that the analyses conducted through the study of Facebook and Instagram platforms relative to cultural sectors

are extremely relevant for three main reasons. Firstly, thanks to the study of the temporal evolution of different metrics regarding the sectors taken into consideration, it will be possible to describe how the cultural sectors studied for inDICEs have used digital tools such as social media. Secondly, through the analyses carried out, it will be possible to outline the impact that communication on social media has had on the fruition of cultural content shared through these platforms. On the basis of these two results, in the further period of analysis it will be possible to compare which types of content are most engaging for a real experience of active participation of online users of the cultural sector.

Regarding chapter 6, "Early results on users behaviour in content co-creation of Wikipedia platform and TikTok social network", the aim of the analyses, conducted by FBK and reported in the present deliverable partially due to an ongoing scientific peer-review process, are to investigate, on one hand, one of the most significant examples of participation and collaboration that we considered with special attention, namely Wikipedia, and its virtual knowledge communities; on the other hand, we aim at exploring TikTok ecosystem in order to understand how creative content re-mix, re-use and production happens in one of the most relevant social media of the current digital sphere.

These analyses are embedded in the investigation and enrichment of one of the most relevant inDICEs theoretical frameworks, namely the theory of the 3.0 Culture, which regards the contemporary strategies and dynamics of cultural production and reproduction in the web 2.0. These analyses can contribute to orient the Policy Guidelines for CHIs and to European Policy Makers for supporting them in the comprehension of state-of-the-art users behaviour, in relation with cultural and creative creation, re-production and the dynamics of virtual community development processes that may pave the way for the evolution of Digital Heritage Communities.

These resources are likely to be useful for a wide range of decision makers, researchers and practitioners in cultural and creative sectors. The reports that

contain aggregated data per CCS will be openly available in the Repository of the inDICEs Observatory Platform.

Our data analyses address the following target groups:

- Policy Makers, who can be informed about the state-of-the-art of the virtual relation between Cultural Heritage Institutions and their users, and be supported in the definition or improvement of policies of active participation via social platforms and in making consequential budget allocation choices;

- Cultural Heritage/Creative Sector Practitioners, who aim at attracting and understanding user experiences for their work, to make use of tools and resources for their professional development, can be supported and guided in better appreciating new ways to spark active participation, to develop digital strategies, tools or practices favouring bottom-up and collective co-creation;

- Researchers and Special Interest Groups searching for relevant data and information on case-studies about the state-of-the-art of the digitization of Cultural Heritage Institutions per sector.

Regarding the originality of the work, this first set of analyses has enriched inDICEs knowledge about digitization processes in the field of CHIs and of cultural production in the web 2.0; literature, at the state-of-the-art, do not provide any equally detailed in depth analysis, over a period of time as long as the one that was here analysed. We can therefore state that the preliminary results brought in this deliverable are central to building a more informed research and development strategy both within the inDICEs project and the European cultural sector, despite being only the beginning of a series of analyses regarding the use of social networks and the study of the digitization of the cultural sector. Finally, the open question regarding the measurability of trends in digital cultural production, consumption and behaviours can be more critically analysed thanks to the present report.

3. Data gathered to Month 24

In the last months of work [M21-24], WP1 kept on collecting data to accomplish the inDICEs objectives of carrying out a close observation of the behaviour and competitiveness of CHIs wanting to be integrated into the Digital Single Market (DSM) and to develop a constant dialogue with their target audiences via digital platforms, including professionals and entrepreneurs from the digital cultural and creative content sectors, in order to implement good practices of digital cultural active participation.

According to the early results that emerged from the first set of data analysis on CHI case studies, we are going to provide some initial suggestions for policy recommendation guidelines.

3.1 Methodology

As reported in D1.3 and D1.4, in order to map the current situation about digital cultural participation of CHIs users, during the first 12 months of the inDICEs project, WP1 gathered a large amount of data from online sources, with special attention to social networks. During the months 12-24 of activity, WP1 proceeded with the building of a first set of case studies (lists of CHIs) per macro-sectoral areas of cultural and creative institutions (namely per CCS), in order to observe the GLAM's organisations and other institutions pertaining to cultural and creative domains as classified by the literature (NESTA, 2006; Thorsby, 2008; Santagata, 2009; KEA, 2019)[1]

The sectors case studies identified and analysed in months 21- 24 are:

- European Fashion GLAMs;

---

[1] NESTA. 2006. Creating Growth How the UK Can Develop World Class Creative Business. NESTA Research Report. NESTA, London; Throsby, D. (2008). Modelling the cultural industries. International journal of cultural policy, 14(3), 217-232.; Santagata, W. (2009). Libro bianco sulla creatività: per un modello italiano di sviluppo. EGEA spa.; KEA new model (2019), see: https://keanet.eu/opinions/culture-nowhere-or-everywhere/ .

- European National Theatres;

- European Archeological Sites.

As already described in D1.4, the lists were drafted by WP1 internal partners and experts; each list was built according to criteria explained in the introduction of each related section, and contains a number between 30 and 50 institutes per sector active in Instagram and/or Facebook. Instagram and Facebook were chosen as the two most widely used and demographically heterogenous social platforms; moreover, the latter is the World most used social platform. According to the "We are social" report (2021)[2], at European level the number of active social media users compared to the total population is between 72% and 79%. Moreover, in both of the two social platforms the percentage of users (ages 16 to 64) is between 75% and 85%. The average age of Facebook users is slightly higher. Data has been collected from July 2019 to July 2021, aggregated per month.

WP1 analyses unpacked data in terms of the following metadata:

- Cultural and Creative Sector;

- Country;

- Month of publication;

- Type of relation between users and CHI as to cultural production impact;

- Form of content sharing (Photos, Links, Statuses, Facebook Videos, Facebook Live, YouTube Videos, Other Videos, Albums, IG Videos, IGTV).

To go into detail of the social platforms taken into account, the data analysis contains the metrics listed below, which are a selected list of social media analytics. In the present deliverable, as reported in the Preliminary Observations section, we considered only the most meaningful metrics that could provide consistent and useful information for the project's objectives.

Instagram:

---

[2] https://wearesocial.com/digital-2021

Account, Codename, URL, Total Interactions, Likes, Comments, Views, All Interaction Rate, Albums, Photos, Video, IGTV, Total Posts, Album Posts, Photo Posts, Video Posts, IGTV Posts, Posts Per Day, Album Likes, Photo Likes, Video Likes, IGTV Likes, Album Comments, Photo Comments, Video Comments, IGTV Comments, Album Video Views, Video Video Views, IGTV Video Views, Followers, Follower Growth, Follower Growth %.

Facebook:

Page, Codename, URL, Total Interactions, Likes, Comments, Shares, Owned Post Views, Owned Views from Shares, Owned Total Views, Percentage Views from Owned Posts, Views on Shared Posts, Views While Live, Video Time, 3-Min+ Videos, Loves, Wows, Hahas, Sads, Angrys, Cares, All Reactions, All Interaction Rate, Photos, Links, Statuses, Facebook Videos, Other Videos, Total Posts, Photo Posts, Link Posts, Status Posts, Owned Video Posts, Shared Video Posts, Other Video Posts, Posts Per Day, Page Likes, Page Growth, Page Growth %, Page Followers, Page Follower Growth, Page Follower Growth %, Interaction Rate Calculated.

The data analysis presented in this deliverable from section 3.2 to section 3.4.2 was conducted by the *Fondazione Bruno Kessler* research team (WP1) using the analytical tool CrowdTangle[3], a content discovery and analytics platform designed to provide content creators with the data and insights they need. CrowdTangle Intelligence gives researchers a way to monitor the performance of a social channel over time, as well as to directly benchmark it against other accounts. Long-term performance figures can help publishers detect overall trends and more easily analyse which content is working and which one is not. CT Intelligence allows users to monitor up to 100 social accounts (on a selection of platforms, that includes Facebook, Instagram, and Reddit) and see overall account-level statistics over time

---

[3] see https://www.crowdtangle.com/

with graphs and charts. It then allows easy comparison of the accounts next to each other, and to export the whole analytics for further use.

3.2 European Fashion GLAMs

This case study is built on the list of the Facebook and Instagram accounts of the European Fashion GLAMs part of the European Fashion Heritage Association (EFHA).

EFHA is an international hub, in which fashion GLAMs (Galleries, Libraries, Archives and Museums) and brands share their digital heritage assets - such as their rich heritage of historical clothing and accessories, contemporary designs, catwalk photographs, drawings, sketches, magazines, catalogues and videos - and their experiences and best practices in the field of digitisation, open access, co-creation, reuse and valorisation of fashion heritage resources, contributing to the digital transformation in the sector and supporting the preservation and valorisation of tangible and intangible heritage connected to costume and, above all, fashion[4]. As noted by Marie Riegels Melchior[5], *valorization* for the EFHA indicates that the initiative is driven by a mission to "actively assert the importance of fashion heritage resources.", allowing the Association to function as a mediator between heritage institutions and private companies; as a disseminator of curated knowledge between members; and as a way for members to share content with site users.

The network of fashion institutions analysed is made of private museums and national institutions of different dimensions and outreach, distributed in 14 European countries. The institutions are united in the objective to unlock and give free access to the unique and vast fashion heritage of Europe. Through their

---

[4] https://fashionheritage.eu/about-us/
[5] Marie Riegels Melchior, 'Digital Fashion Heritage. Understanding europeanafashion.eu and Google Cultural Institute', *Critical Studies in Fashion & Beauty*, Volume 10 Number 1.

collections, they aim at finding a better-suiting definition of European identity, acknowledging the past and looking at better ways to understand the present and shape the future; they also want to unlock the full potential of their shared fashion heritage for creatives, scholars and fashion lovers alike to admire and reactivate through their practices, always in an open and careful spirit. In addition, through the valorisation of their shared fashion heritage for work, for study and for fun, they foster a more inclusive, diverse and creative society and support the preservation and valorisation of craftsmanship and the related intangible heritage connected to fashion and all its outcomes, to inspire and foster a more sustainable and fair future for the whole sector[6].

We tried to maintain the samples used for the Facebook and Instagram platforms as close as possible, provided that not all the museums selected had both comparably active Instagram and Facebook pages that could be amenable to data analysis.

The list of organisations considered in the analysis is composed as follows:

CatwalkPictures, Modemuseum Hasselt, Mode Museum Antwerpen (MOMU), Musée Mode & Dentelle, Palais Galliera - Musée de la mode, Musée Yves Saint Laurent, Cité de la dentelle et de la mode, Museo della Calzatura di Villa Foscarini Rossi, Fondazione Gianfranco Ferré, Fondazione FILA Museum, Fondazione Cerratelli, International Talent Support, Armani/Silos, Museo del Tessuto di Prato, Salvatore Ferragamo Museum, Modemuze, Textiel Museum, Textile Research Centre, Centralne Muzeum Włókiennictwa w Łodzi, Museu Do Design e Da Moda (MUDE), Museo del Traje, Cristóbal Balenciaga Museoa, Cristóbal Balenciaga Museoa, Fashion Museum Bath, Fashion and Textile Museum, Westminster Menswear Archive, Fondazione Zegna, Yorkshire Fashion Archive, Pucci Archive, Archivio Missoni.

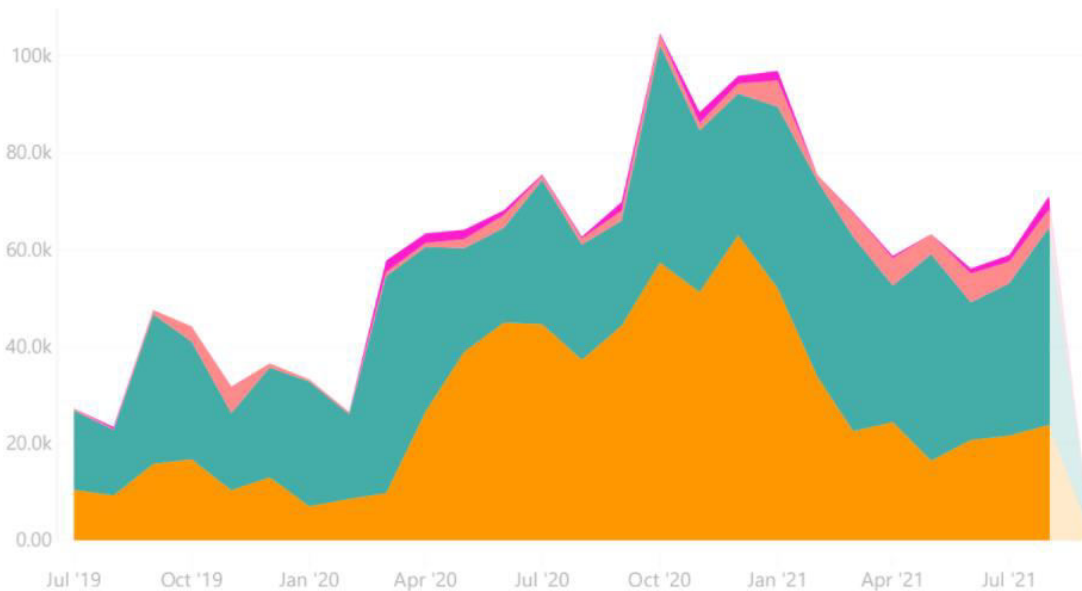[6] https://fashionheritage.eu/about-us/

3.2.1 Instagram

Follower Growth [figure 1 below]: This metric measures the number of new followers that the list of the entire case study of Fashion GLAMs gained on Instagram over that set period of time. This provides an indication of the "share of conversation" captured and consequently of the success on a certain platform, but not of the level of active participation of the users. The percentage of Instagram followers of the list of Fashion Institutes increased by 86% in two years.

FOLLOWER GROWTH   >   07/01/19 - 09/01/21

	FOLLOWERS	GROWTH	% GROWTH
>	396.7K	+184.5K	+86.90%

Total interactions [figure 2 below]: Total interactions represent the sum of different social media actions, namely reactions, comments and shares. Interactions are also known as engagement, and represent a metric that can reveal an active response of the users if positively compared to a growing line graph regarding interaction rate. Also from this point of view, the trend reveals a main peak, namely in the autumn

and winter of 20/21, which corresponds to the second European lockdown period. Photos are way up the most engaging type of content, followed by albums while the tool of IGTV seems very marginal.

ALL POSTS   >   TOTAL INTERACTIONS   >   07/01/19 - 09/01/21

	TOTAL INTERACTIONS	AVG. MONTHLY INTERACTIONS	MONTHLY INTERACTION RATE
>	1.58M	58.1K	4.52%

Post Type	Total	%	
Albums	726.9K	89.48%	
Videos	63.3K	7.80%	
IGTV	22.1K	2.72%	

Interaction rate [figure 3 below]: The Interaction Rate is calculated by adding up all the interactions on every post from every account in the list (weighted by its historical records), and then dividing it by the total number of posts and by the average size (follower count/page likes) of the respective account. Here too we can find various peaks, the most important of which is observed in the spring of 2020, in correspondence with the first lockdown. In general, we can observe how the interaction rate strongly decreased in the last year, while before the first lockdown period (Spring 2020) we can notice 4 picks of interactions.

15
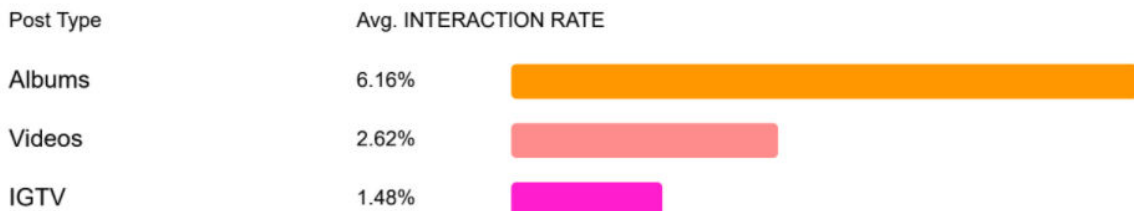
Total views [figure 4 below]: Also in this case there are different peaks of activity, mainly concentrated around the three lockdowns, but one can also observe that the highest peak corresponds to the period in which the world of Fashion restarted to host and produce live events. The highest increase of views here is peculiar and it is located in Summer 2021. Videos and IGTV are the two main typologies of contents visualised, but videos, which are more likely to be pre-recorded and with edited contents, gain more visualisations then IGTV, a tool more likely to be live-streamed.
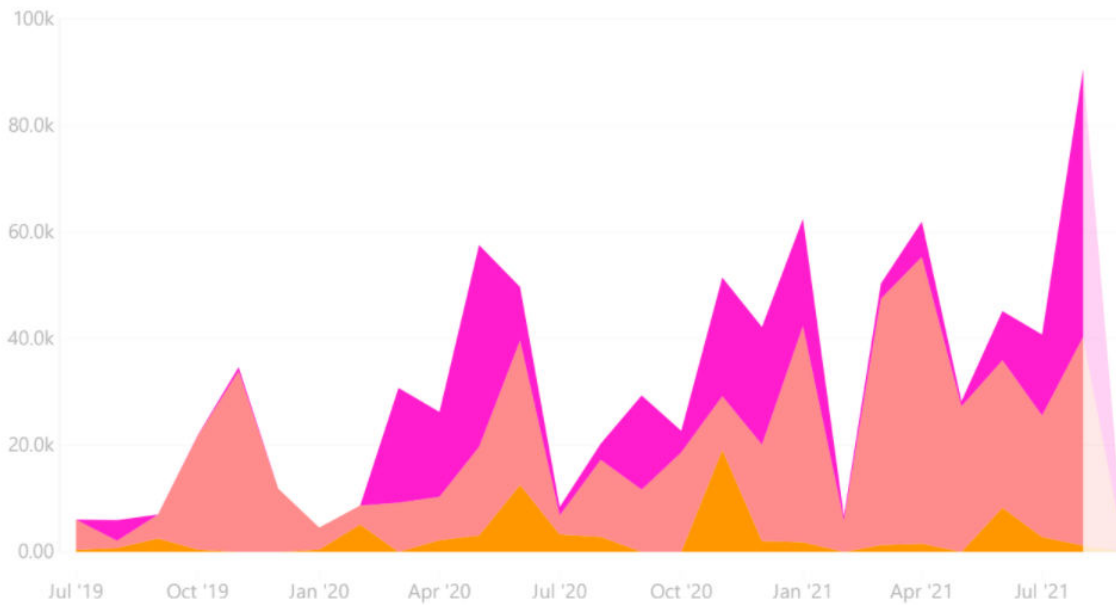
ALL POSTS > TOTAL VIEWS > 07/01/19 - 09/01/21

|	TOTAL VIEWS	TOTAL POSTS
	826.7K	1,259

Post Type	# Views	%	# Posts	
Albums	71.7K	8.67%	948	
Videos	488.4K	59.08%	188	
IGTV	266.6K	32.25%	123	

3.2.2 Facebook

Follower Growth [figure 5 below]: This metric measures the number of new followers that the entire case study list of museums gained on Facebook over the set period of time. The percentage of Facebook followers of the list of Fashion Institutes increased by 18,6% in two years. In the summer of 2020 (September 2020 +30.1K), we can observe the highest growth during the covid-19 crisis. From winter 2020, we can observe a constant but slow growth.

FOLLOWER GROWTH > 07/01/19 - 09/01/21

FOLLOWERS	GROWTH	% GROWTH
824.6K	+129.3K	+18.60%

Total interactions [figure 6 below]: On Facebook, the type of content that triggers most reactions and engagement so far is "Photos".
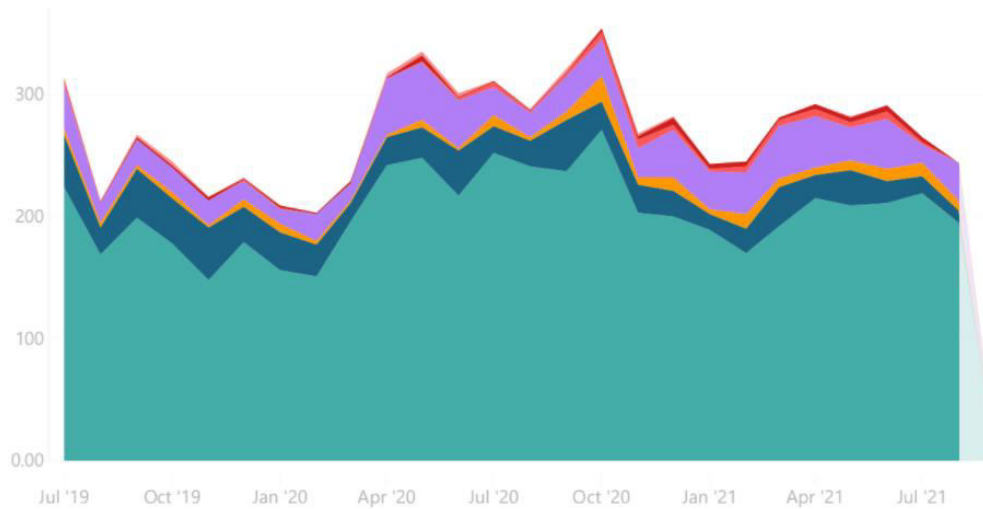
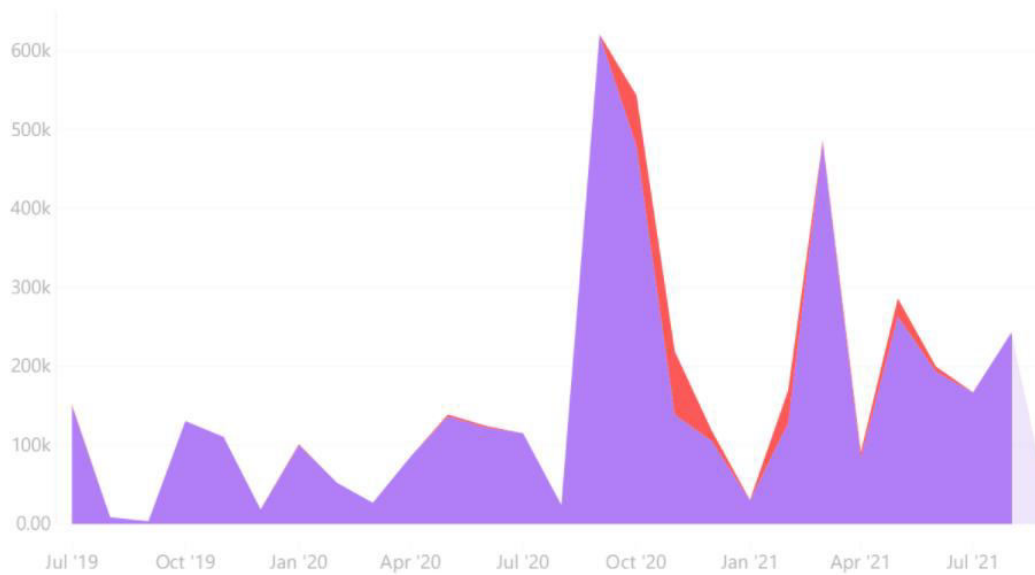| Post Type	Total	%
Photos	5.3K	75.36%
Links	679	9.62%
Statuses	170	2.41%
Facebook Videos	739	10.47%
Facebook Live	63	0.89%
YouTube Videos	64	0.91%
Other Videos	24	0.34%

Interaction rate [figure 7 below]: We can observe how the interaction rate undergoes strong fluctuations but, in general, it seems that Fashion Institutes cyclically retain the attention of their users by stimulating their active participation.
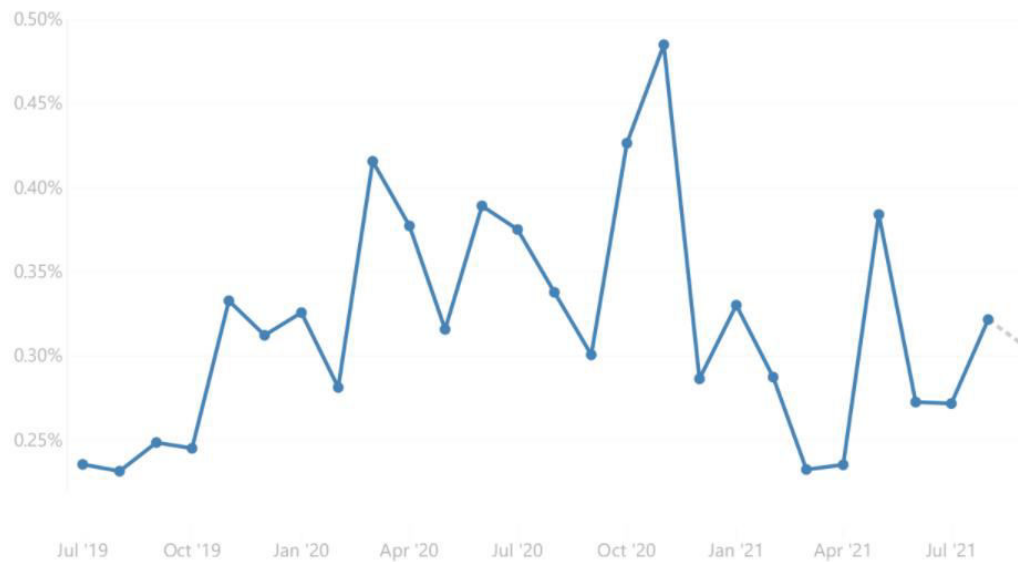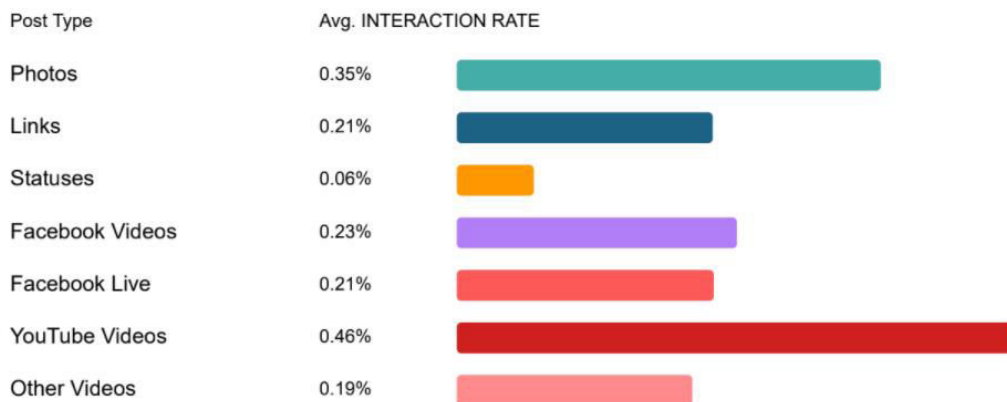
Total views [figure 8 below]: Videos created and shared on Facebook pages from the selected museums of this case study seem to be a very effective channel for reaching out to digital users, but the trend of the graph line is discordant with that of the interaction (engagement) rate.

Post Type	# Views	%	# Videos	
Facebook Videos	4.03M	94.22%	675	
Facebook Live	247.3K	5.78%	53	

3.3 European National Theatres

The case study list of the National Theatres in Europe is built on the analysis of the members of networks of theatres working together in Europe.As for the Fashion GLAMs sector, the performing live sector reaches its peak of expression and happens ontologically in the same moment of its "performance", namely when it is Live Theater. The same digital reproduction of the live performance is still not homogeneously considered as "pure" theatre, but trespasses and overlaps with the borders of the film/audio-visual sector.

The main objectives of the data collection and analysis of this case study are the observation of the initiatives of the selected Theatres and of the responses of their

users in the pandemic period compared to the previous year, in order to detect current trends in terms of digital participation in the sector. To select the sample, we considered 46 National Theatres pertaining to 24 countries situated in the whole European territory. We focused on "National" Theatres as they are essentially producers: they produce alone, or in collaboration with others, shows that are often circulated on tour or rented, and can be differentiated from the "cultural" Theatres, which mostly produce some shows on their own, but their main activity is to propose and present installations produced elsewhere. We tried to maintain the samples used for Facebook and Instagram as similar as possible, given that not all the selected museums had both comparably active Instagram and Facebook pages that could be amenable to data analysis.

The list considered in the case study is as follows:

Teatri Kombetar, Landesthater Linz, Schauspielhaus Graz, Théâtre de Liège, ТЕАТРАЛНА РАБОТИЛНИЦА СФУМАТО, ТЕАТРАЛНО-МУЗИКАЛЕН ЦЕНТЪР, Hrvatsko Narodno Kazaliste, Národní divadlo, Městská divadla pražská, Θεατρικού Οργανισμού Κύπρου, Suomen Kansallisteatteri, CDN Nancy Lorraine - La Manufacture, Le théâtre de l'Odéon, Kote Marjanishvili State Drama Theatre, Kutaisi Lado Meskhishvili Professional State Drama Theatre, Feutsches Theater Berlin, Theater und Orchester Heidelberg, Κρατικό Θέατρο Βορείου Ελλάδος, Pesti Magyar Színház, Weöres Sándor Színház, Fondazione Teatro Due, Teatro Koreja, LATVIJAS JAUNĀ TEĀTRA INSTITŪTS, Lietuvos Nacionalinis Dramos Teatras, Les théâtres de la ville du Luxembourg, Escher Theatre, Teatru Malta, Det norske theatret, JK Opole Theatre, Narodowy Stary Theatr Krakow, Teatro Nacional Sao Joao, Teatro Nacional D. Maria II, Teatrul national timisoara, Teatrul Alexandru Davila, Narodno Pozoriste, Jugoslovensko Dramsko Pozorište, Slovenské národné divadlo, Ján Palarika's Theater, Prešernovo Gledališče Kranj, Slovensko Narodno Gledalisce Nova Gorica, Göteborgs Stadsteater, De Toneelmakerij, Dakh Theatre - Centre of Contemporary Arts - "dakh", Kyiv National Academic Molodyy Theatre, Belarus Free Theatre.

3.3.1 Instagram

Follower Growth [figure 9 below]: The number of new followers that the entire case study list of National Theatres gained on Instagram over the last two years continuously increased by more than 163%. In two time windows corresponding to the first lockdown period and winter 20-21 (around Christmas - New Year period), we can observe the highest growth. However, we can observe positive growth throughout.
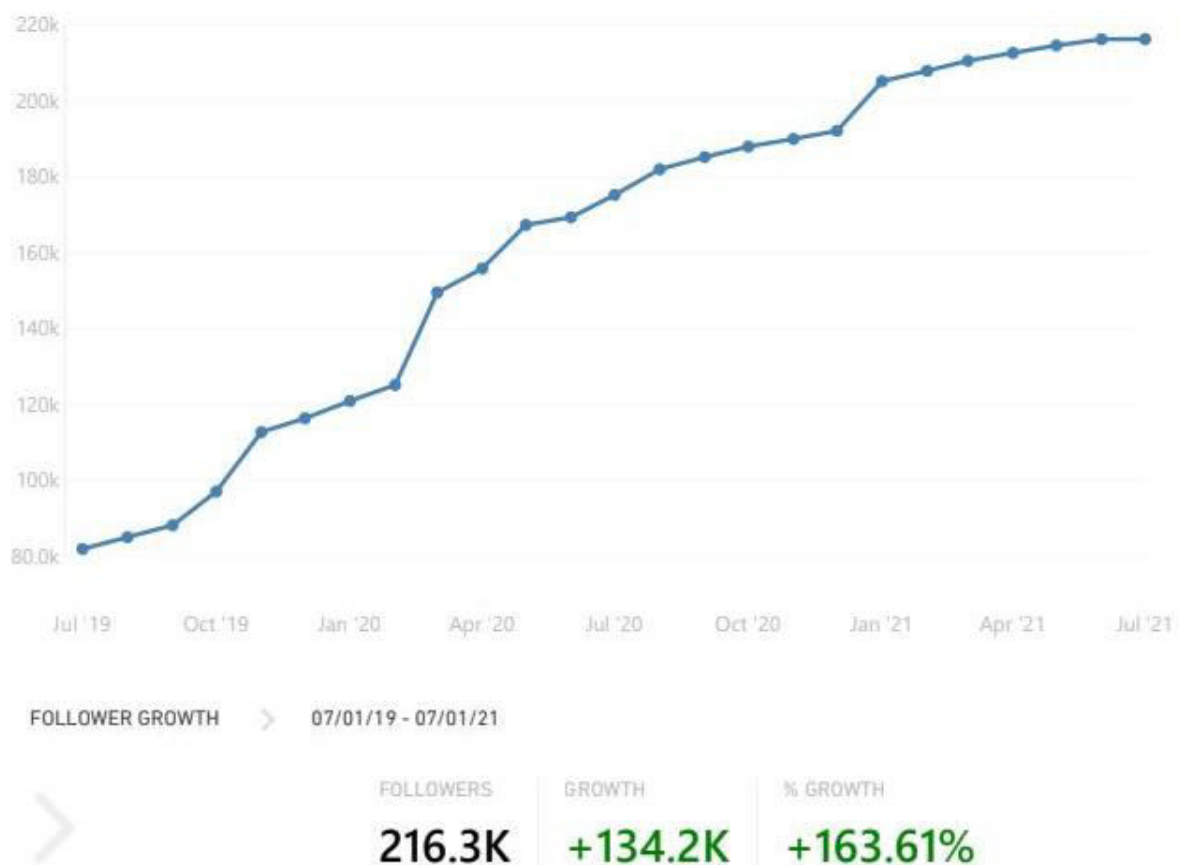

Total interactions [figure 10 below]: The trend reveals various peaks that cyclically grow and decrease, highlighting waves of renewal of interest toward, and interaction with, the selected theatres by their digital visitors. Photos and albums are the type of content that generate the most interactions.

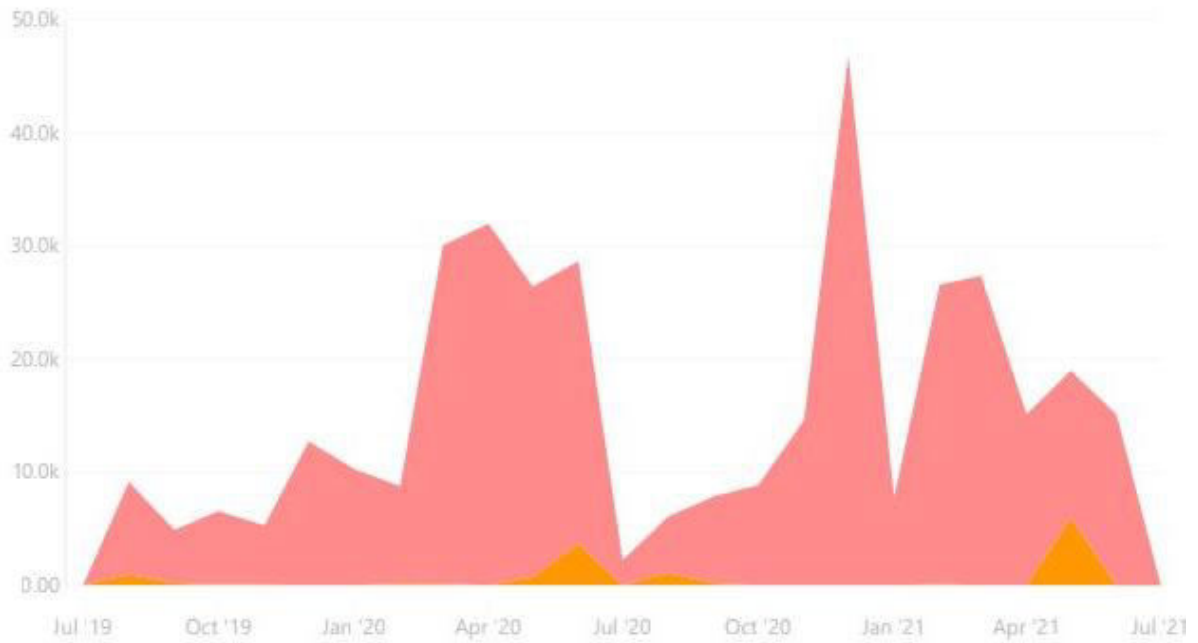| Post Type	Total	%	
Albums	375.2K	25.64%	
Photos	901.7K	61.62%	
Videos	186.6K	12.75%	
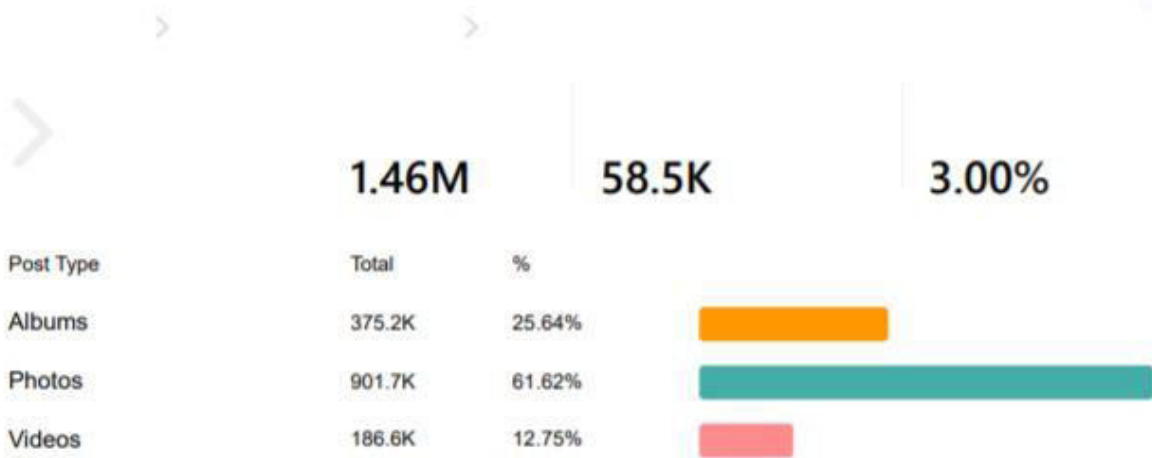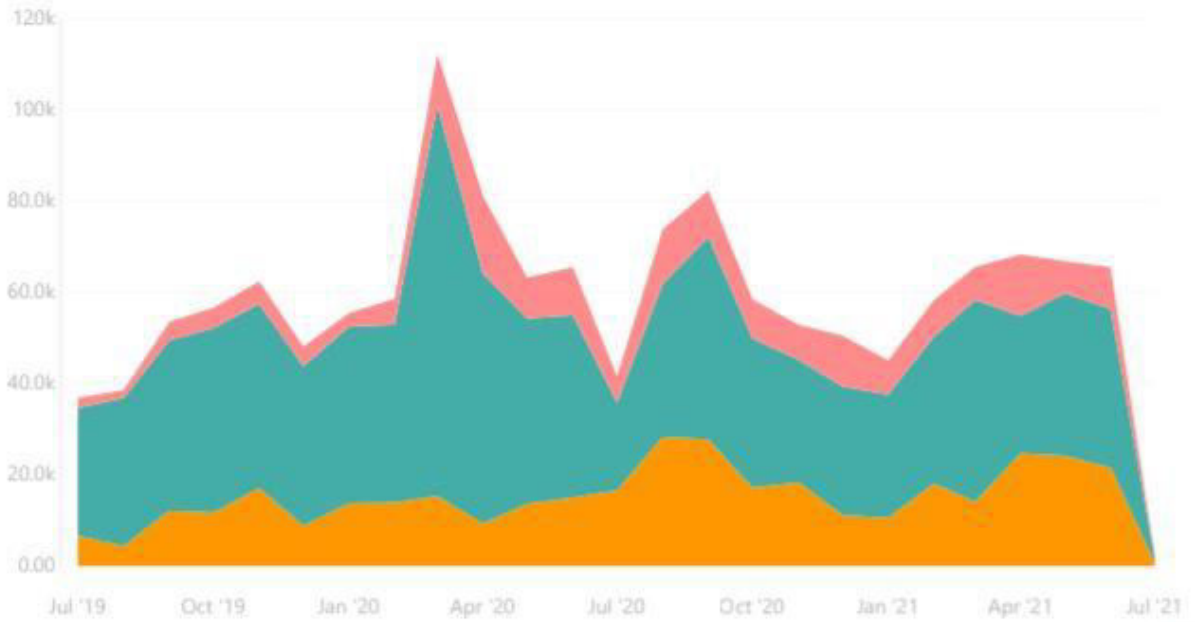
Interaction rate [figure 11 below]: The Interaction Rate is here characterised by various peaks of renewed waves of active participation by visitors but, in general, we can observe how the interaction rate decreased in the last two years overall. IGTV isn't even considered among the tools for sharing contents.
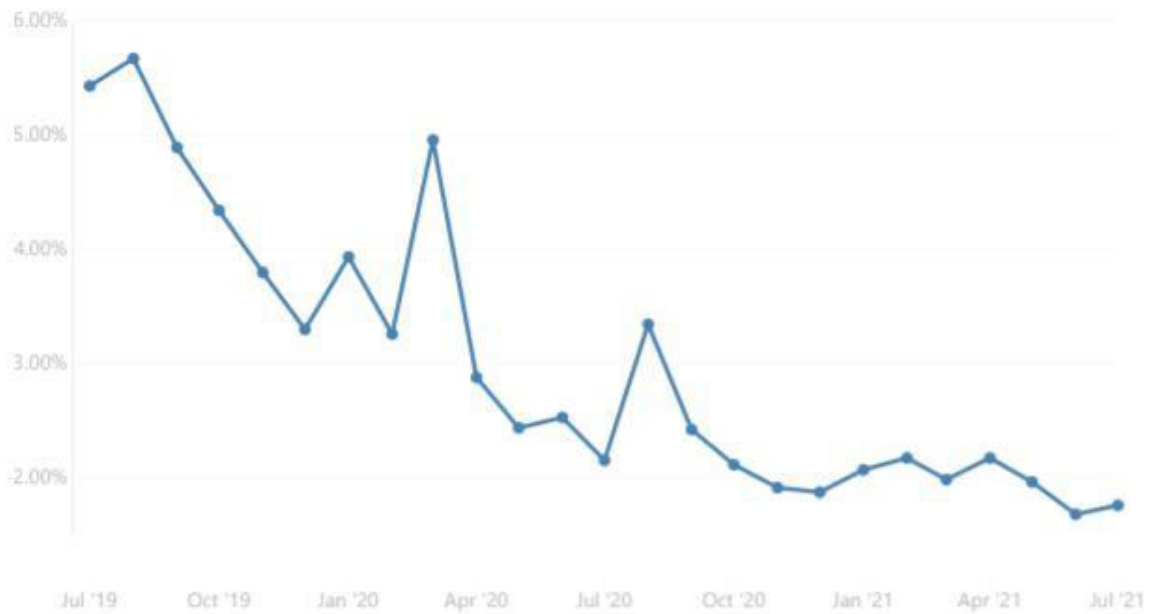
Total views [figure 12 below]: Video here is the most viewed channel of content sharing, but IGTV is not even considered both by the National Theatres pages and (consequently) by their users. Also here we can find different peaks, but in the time window related to the winter of 2020 and 2021 we once more find the highest increase of views.
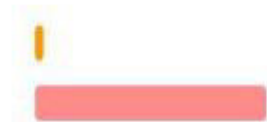
3.3.2 Facebook

Follower Growth [figure 13 below]: The percentage of Facebook followers of the listed National Theatres grew around 46% in the last two years, with a high peak

around   the   end   of   2019.
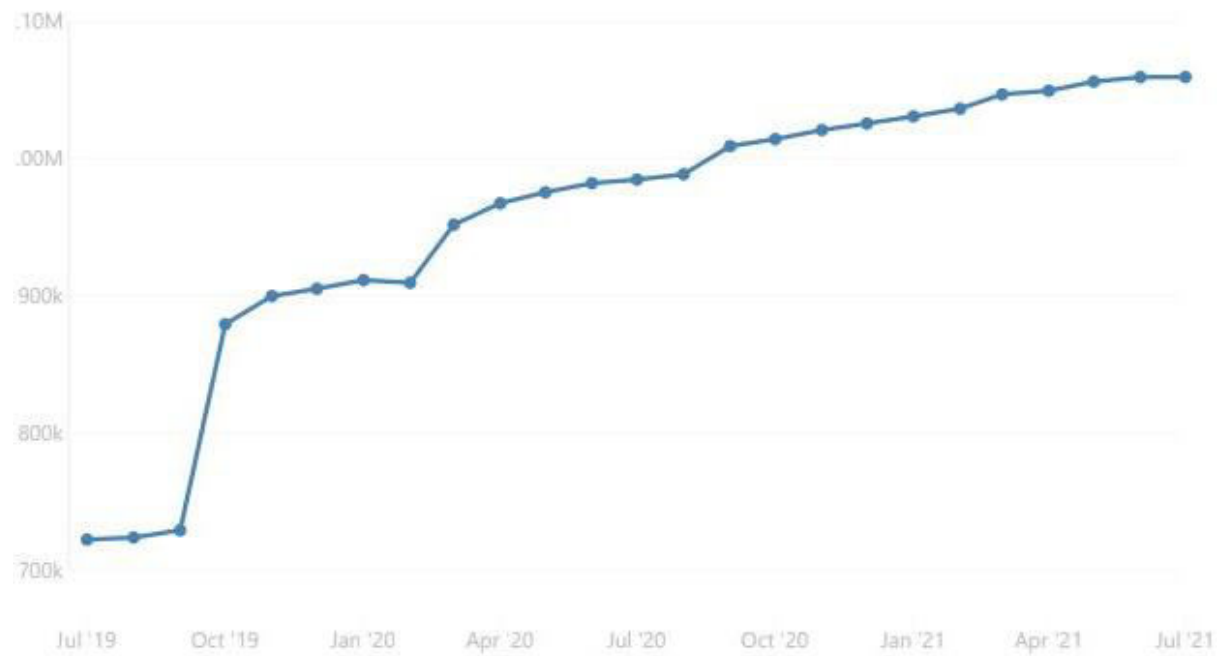
FOLLOWER GROWTH   >   07/01/19 - 07/01/21

	FOLLOWERS	GROWTH	% GROWTH
>	1.06M	+337.0K	+46.64%

Total interactions [figure 14 below]: The trend reveals a main peak in the spring of 2020, which once more corresponds to the first lockdown period. On Facebook, the types of content that trigger the most reactions and engagement are "Photos" and "Videos".
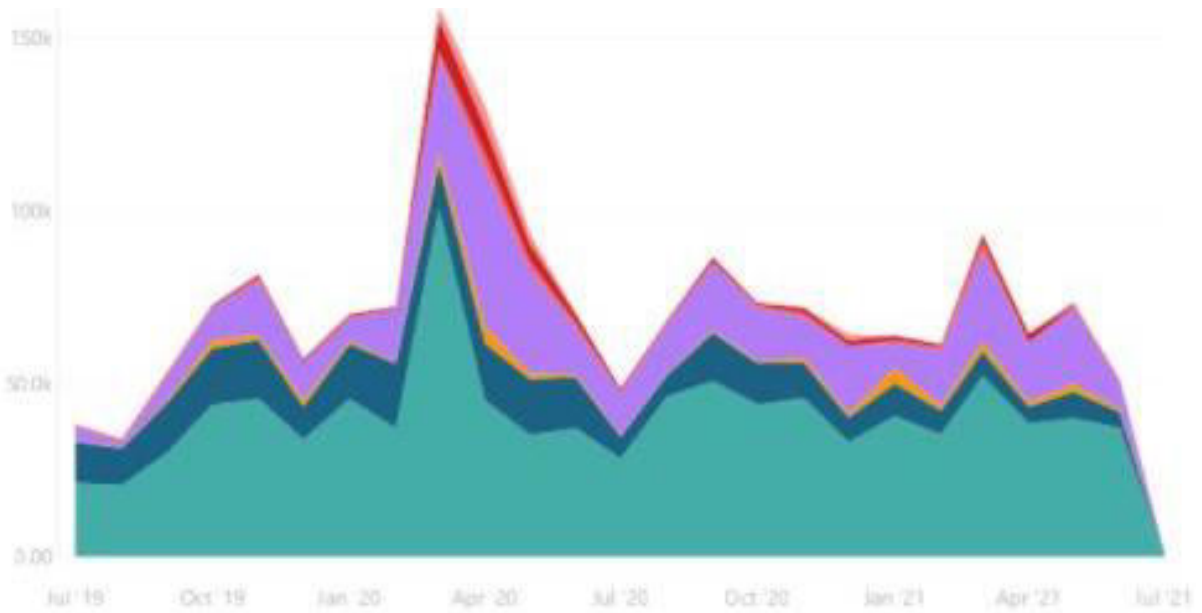
ALL POSTS   >   TOTAL INTERACTIONS   >   07/01/19 - 07/01/21

	TOTAL INTERACTIONS	AVG. MONTHLY INTERACTIONS	MONTHLY INTERACTION RATE
	1.76M	70.3K	0.35%

Post Type	Total	%	
Photos	990.4K	56.37%	
Links	271.3K	15.44%	
Statuses	36.1K	2.05%	
Facebook Videos	379.5K	21.60%	
Facebook Live	19.5K	1.11%	
YouTube Videos	42.3K	2.41%	
Other Videos	17.9K	1.02%	

Interaction rate [figure 15 below]: The interaction rate percentage is higher here regarding Facebook Videos and Live streaming, with various peaks all along 2020.

28
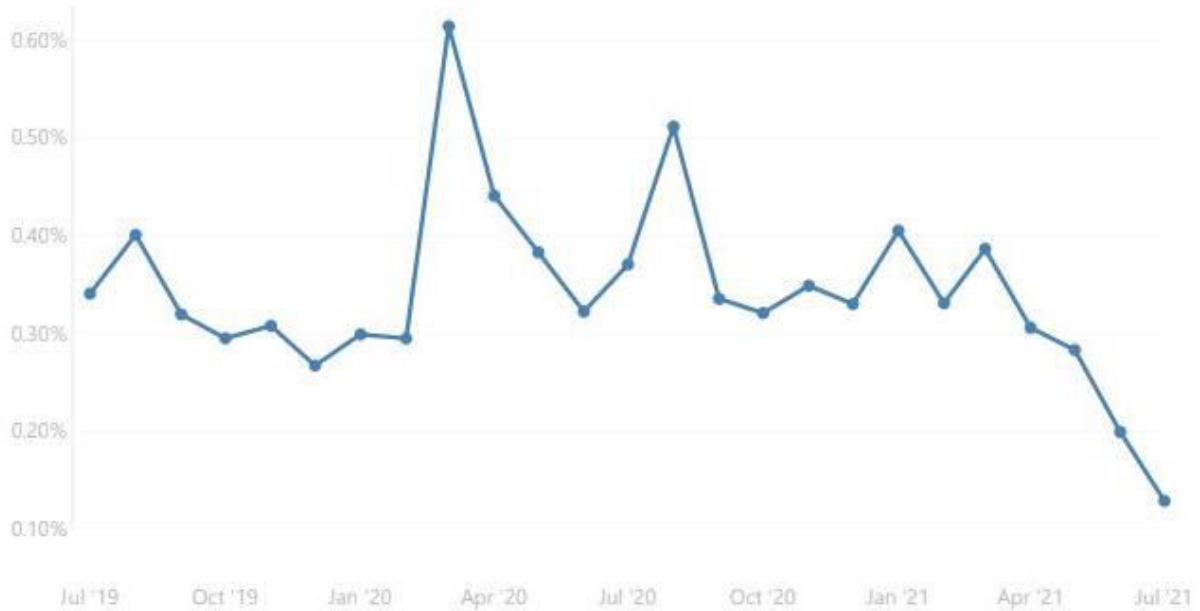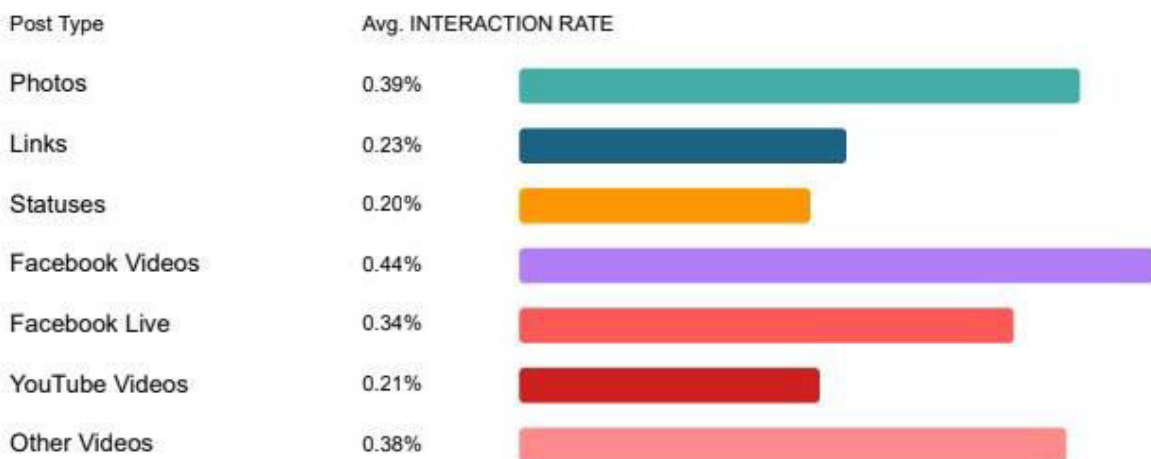
Total views [figure 16 below]: In the last two years posted videos have been of course the 96%, with various very high peaks corresponding mainly with the first lockdown, given that most of the Live Videos couldn't have been produced for the lack of live performances.

3.4 European Archeological Sites

This case study is built on the list of the Facebook and Instagram accounts of the European Archeological Sites. The peculiarities of this list regard the sample and the non-homogeneity of the countries that host Archeological Sites with an Instagram or Facebook page. Indeed, the sample is quite small when compared to the other sectors, and the geographic distribution of the sites is uneven covering the

following percentages: Italy and France 17.6%, United Kingdom and Ireland 11.8%, 5,9% for Greece, Lithuania, Austria, Germany, Portugal, Spain, Czech Republic.
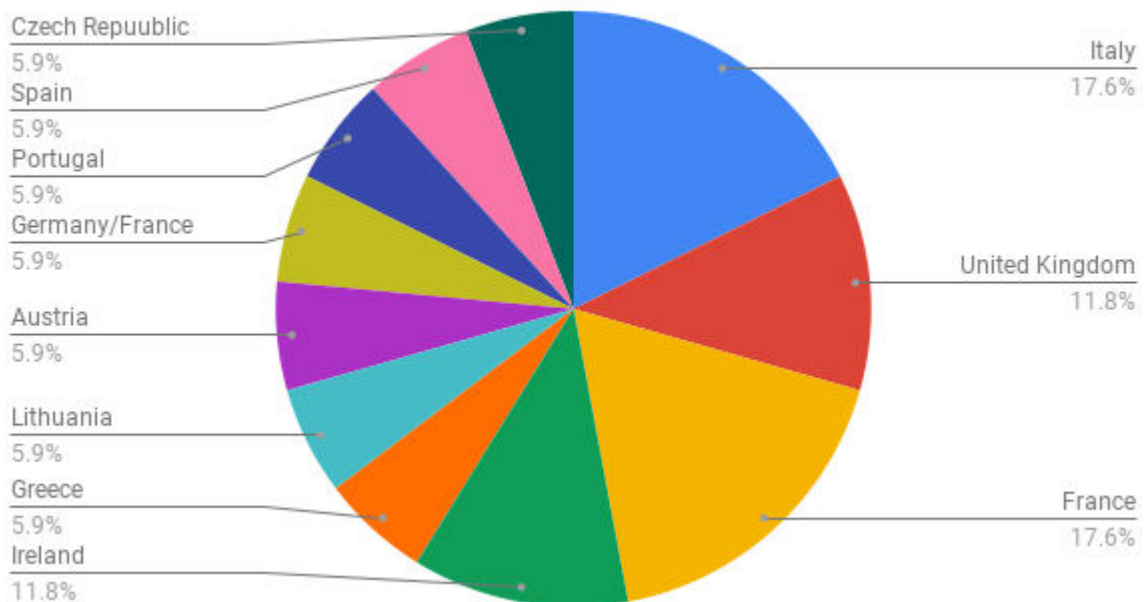
In general, the sample is highly influenced by the low level of digitization of National Archeological CHIs. Many outstanding European archaeological sites – even included in the UNESCO list - don't have their own website or online profile and/or don't use social media. Moreover, many public ones are under the umbrella of Ministries of Culture, heritage national agencies or local municipalities websites.

For what we could observe by a direct observation of the Instagram and Facebook pages of this list, most of the content is directly created by users at the moment when they are visiting the site, by posting pictures, videos, live videos or by geolocalizing themselves.

The list covered in the case study is as follows:

Parco archeologico del Colosseo, Stonehenge, Pont du Gard, The Roman baths of Bath, Brú na Bóinne, Arènes de Nîmes, Parco archeologico della Valle dei templi di Agrigento, Αρχαιολογικός Χώρος Δελφών, Skellig Michael, Kernavės archeologinė

vietovė, Römerstadt Carnuntum, European Archaeological Park at Bliesbruck-Reinheim, Museo e parque archeologico do Vale do Côa, Empúries, Parco archeologico di Pompei, Hrad Helfštýn, Théâtre Antique & Musée d'Orange.

3.4.1 Instagram

Follower Growth [figure 17 below] The trend reveals a 97% growth in the last two years.
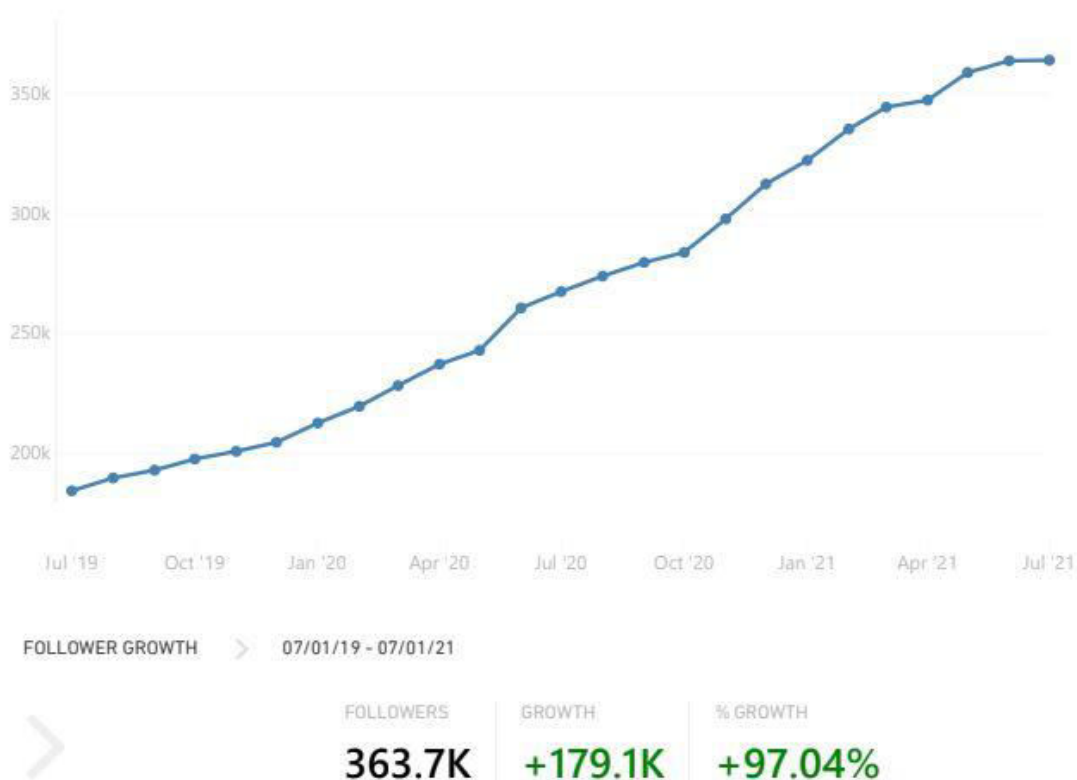
Total Interactions [figure 18 below] "Photo" is the most important and engaging type of content. The trend shows a big increase in the interactions during 2021, with a peak that begins to mount in the fall of 2020, namely during the second lockdown period.
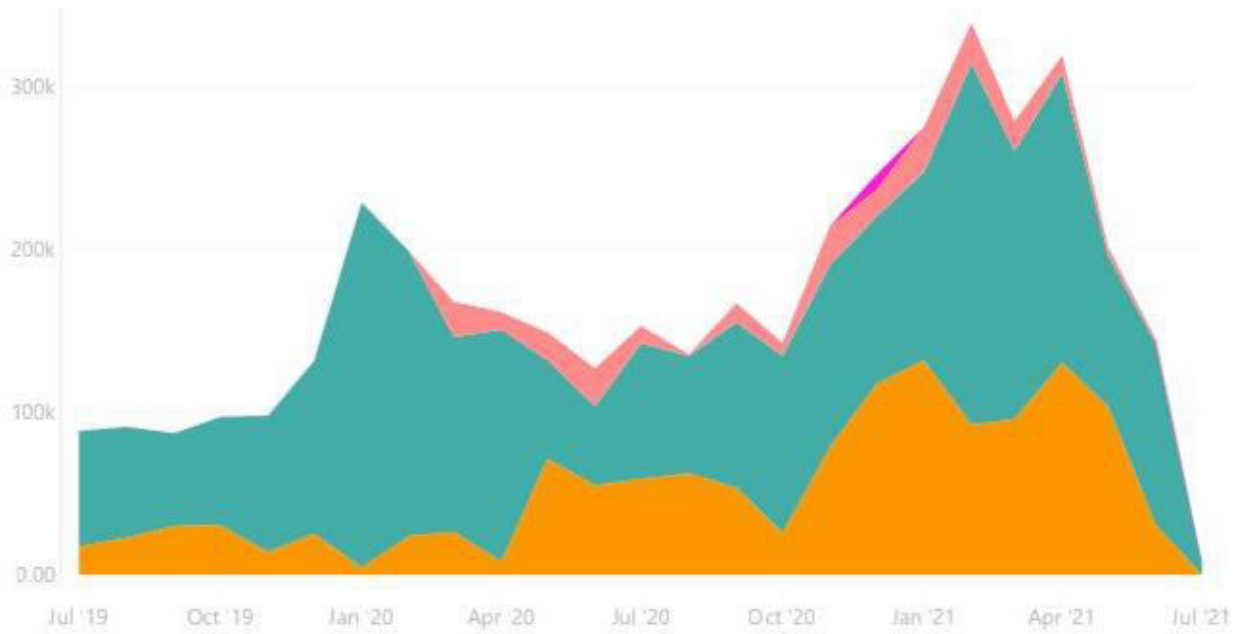
ALL POSTS > TOTAL INTERACTIONS > 07/01/19 - 07/01/21

	TOTAL INTERACTIONS	AVG. MONTHLY INTERACTIONS	MONTHLY INTERACTION RATE
>	**4.24M**	**169.8K**	**8.99%**

Post Type	Total	%	
Albums	1.31M	30.83%	
Photos	2.69M	63.44%	
Videos	233.7K	5.51%	
IGTV	9.5K	0.22%	

Interaction rate [figure 19 below]: The Interaction Rate is calculated by adding up all the interactions on every post from every account in the list (weighted by its historical records), and then dividing it by the number of posts and by the average size. As the previous graph shows, while content posted increased, the interaction rate diminished over time, revealing a decline in user interest, even if the line presents some peaks. The most used channel of content sharing is IGTV.
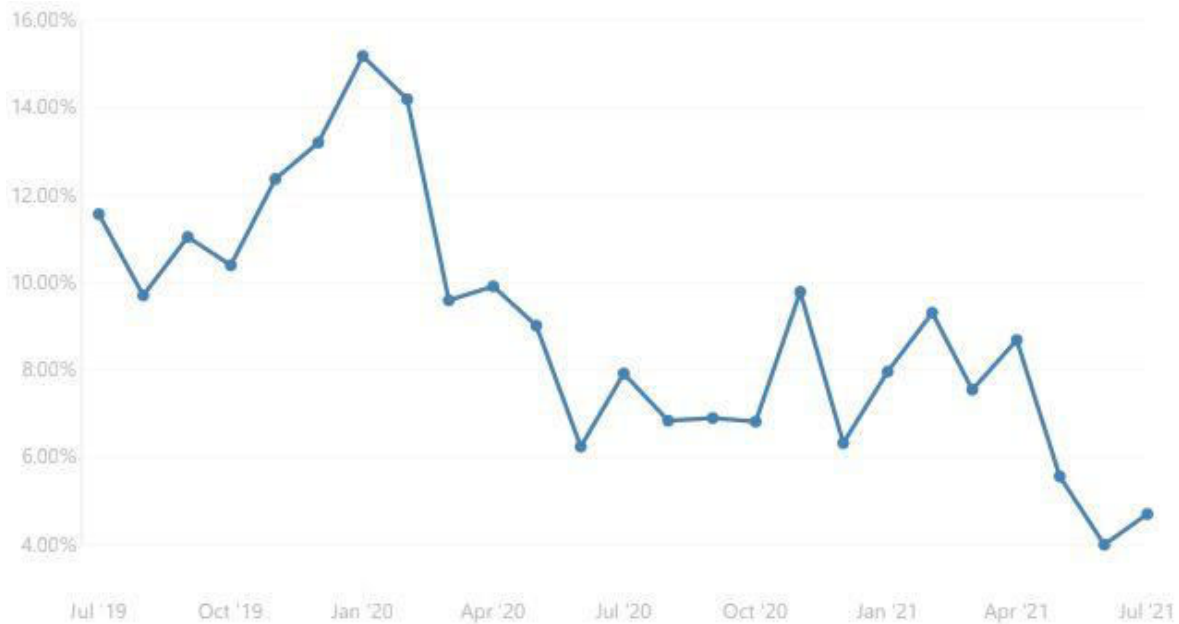
ALL POSTS > AVG. INTERACTION RATE > 07/01/19 – 07/01/21

	MONTHLY INTERACTION RATE	AVG. MONTHLY INTERACTIONS	TOTAL POSTS
>	8.99%	169.8K	2.6K

Post Type	Avg. INTERACTION RATE	
Albums	16.45%	
Photos	7.76%	
Videos	4.28%	
IGTV	42.48%	

34
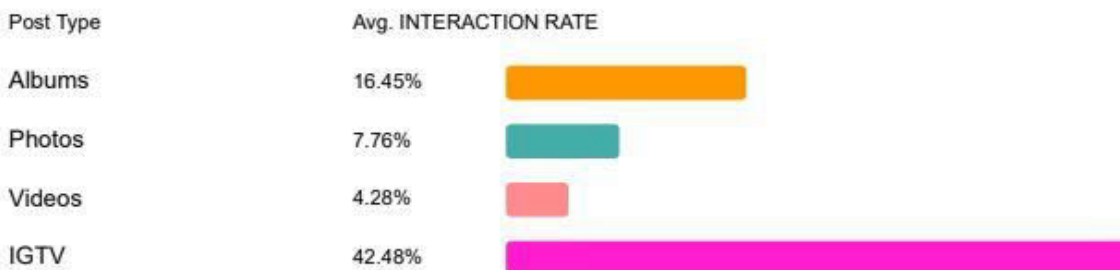
Total views [figure 20 below]: Videos are the most viewed content, with around 2 millions views (both videos and IGTV) per only around 600 posts published in two years per category. Also here we can find three different peaks: spring 2020, fall 2020 and winter 2021: what we can observe is that, before the first wave of pandemic, European Archeological Sites Instagram pages didn't even consider the Video as a tool for sharing contents: indeed, they started to produce video content around the end of February 2020.
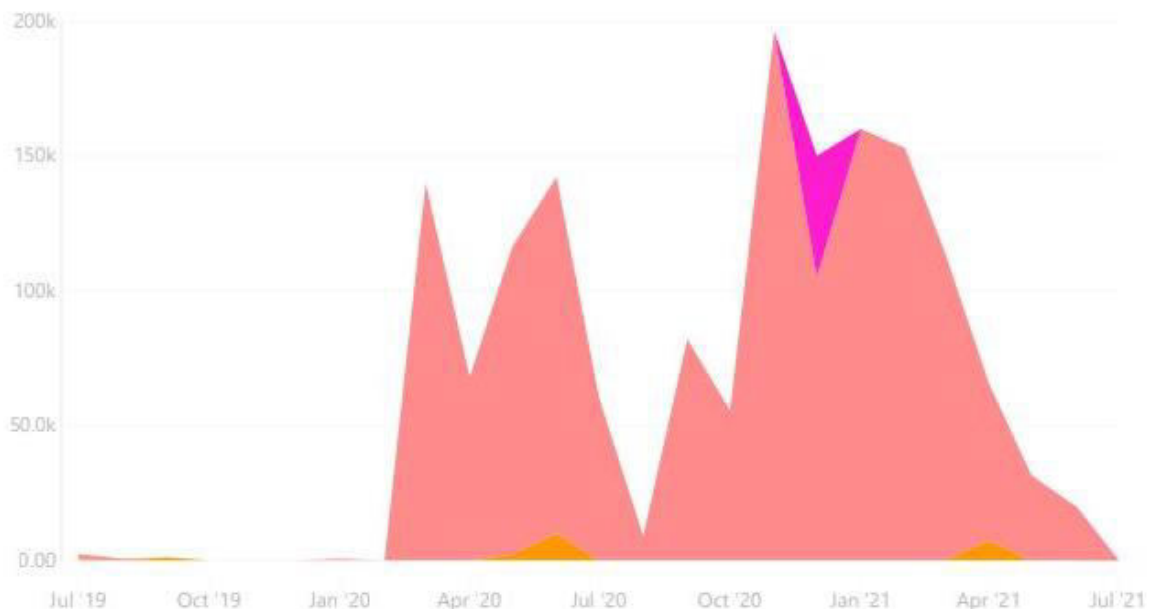
ALL POSTS > TOTAL VIEWS > 07/01/19 - 07/01/21

	TOTAL VIEWS	TOTAL POSTS
>	**1.57M**	**629**

Post Type	# Views	%	# Posts	
Albums	21.1K	1.34%	400	
Videos	1.50M	95.79%	228	
IGTV	45.1K	2.87%	1	

3.4.2 Facebook

Follower Growth [figure 21 below]: The trend reveals a constant growth that goes up to + 86% in the last two years.
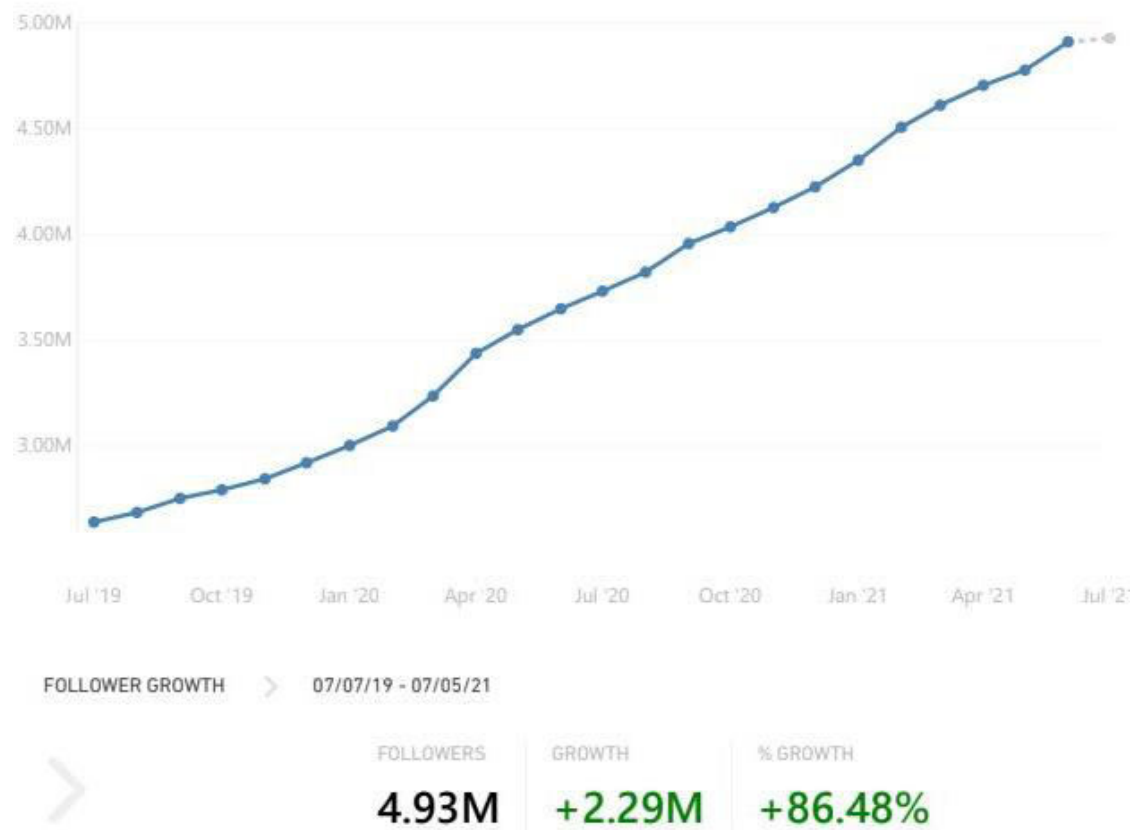
Total interactions [figure 22 below] Total interactions represent the sum of different social media actions, such as likes, comments or reactions to a post. Also from this point of view, the trend reveals three different peaks, namely the spring and fall of 2020 and winter 2021. Video is by far the most engaging type of content, with a total of more than 41 millions of interactions in the last two years.
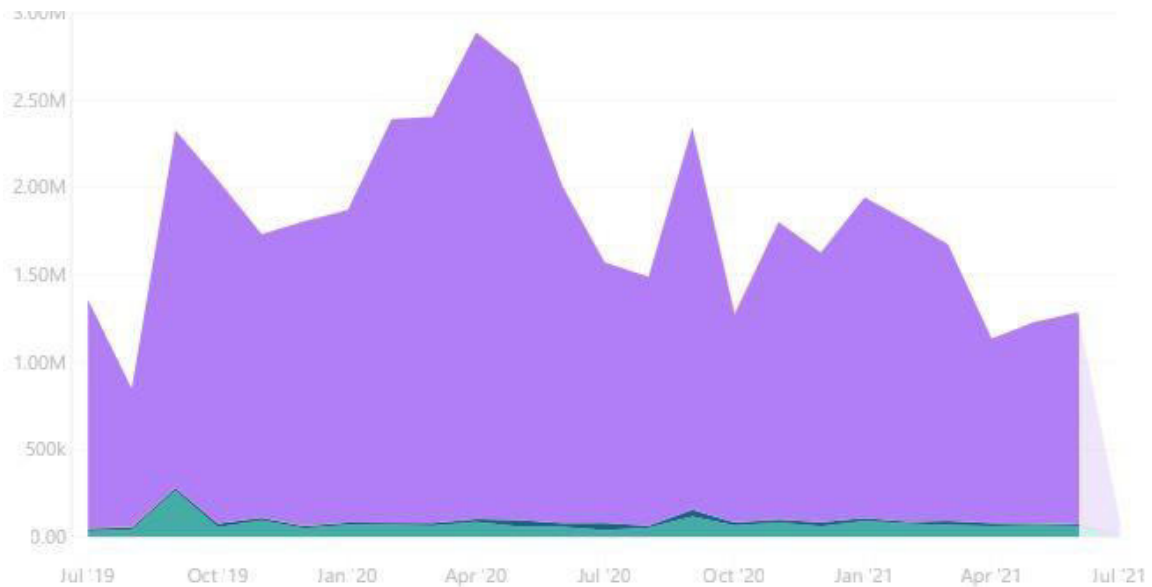
ALL POSTS > TOTAL INTERACTIONS > 07/07/19 - 07/05/21

|	TOTAL INTERACTIONS	AVG. MONTHLY INTERACTIONS	MONTHLY INTERACTION RATE
>	**43.66M**	**1.69M**	**1.37%**

Post Type	Total	%	
Photos	1.80M	4.12%	
Links	437.4K	1.00%	
Statuses	22.5K	0.05%	
Facebook Videos	41.36M	94.75%	
Facebook Live	5.3K	0.01%	
YouTube Videos	16.6K	0.04%	

Post count [figure 23 below]: The chart represents a simple count of how many posts have been produced by the whole list accounts per each tool of content production. As we can see, the time series shows different constant peaks in posting videos, links and photos maily. This pattern cannot therefore merely be attributed to changes in users' choices related to the new situation created by the pandemic.
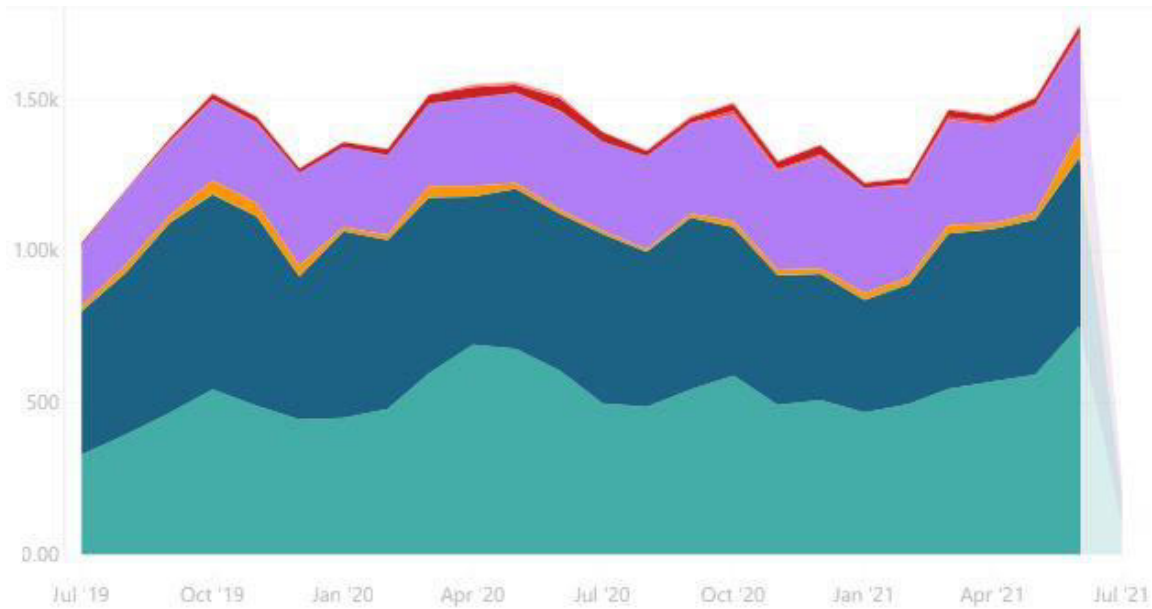
ALL POSTS > TOTAL POSTS > 07/07/19 - 07/05/21

	TOTAL POSTS	AVG. MONTHLY POSTS
>	33.9K	1,311

Post Type	Total	%	
Photos	12.8K	37.81%	
Links	12.6K	37.09%	
Statuses	622	1.83%	
Facebook Videos	7.2K	21.23%	
Facebook Live	97	0.29%	
YouTube Videos	508	1.50%	
Other Videos	87	0.26%	

Interaction rate [figure 24 below] The graph line is pretty erratic for the whole two years. It mainly refers to Facebook videos and, in general, it decreases.

38
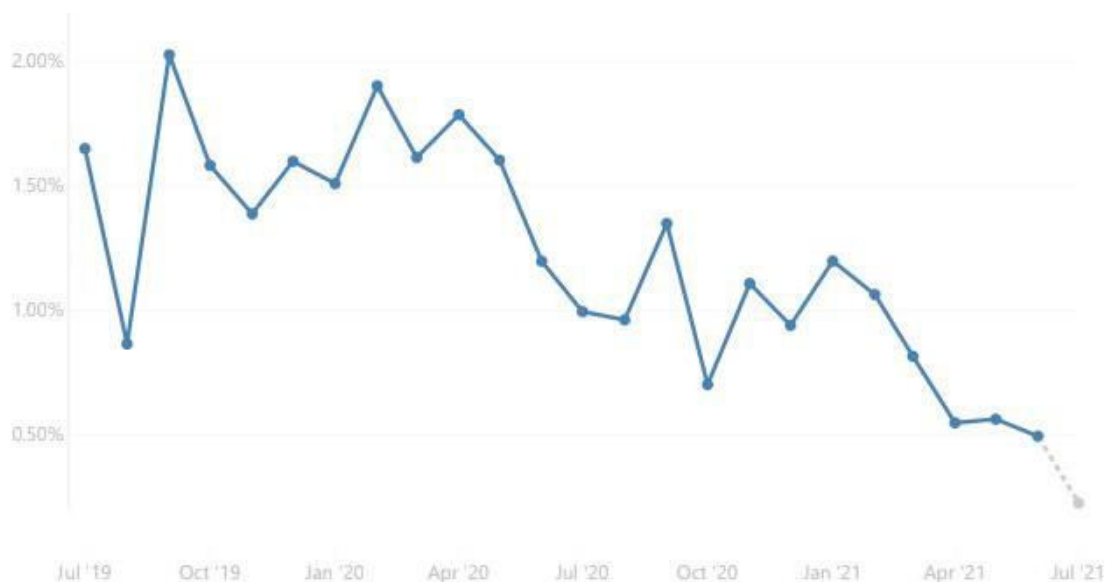
ALL POSTS > AVG. INTERACTION RATE > 07/07/19 - 07/05/21

	MONTHLY INTERACTION RATE	AVG. MONTHLY INTERACTIONS	TOTAL POSTS
>	1.15%	1.69M	33.9K

Post Type	Avg. INTERACTION RATE	
Photos	0.14%	
Links	0.03%	
Statuses	0.04%	
Facebook Videos	5.50%	
Facebook Live	0.05%	
YouTube Videos	0.03%	
Other Videos	0.11%	
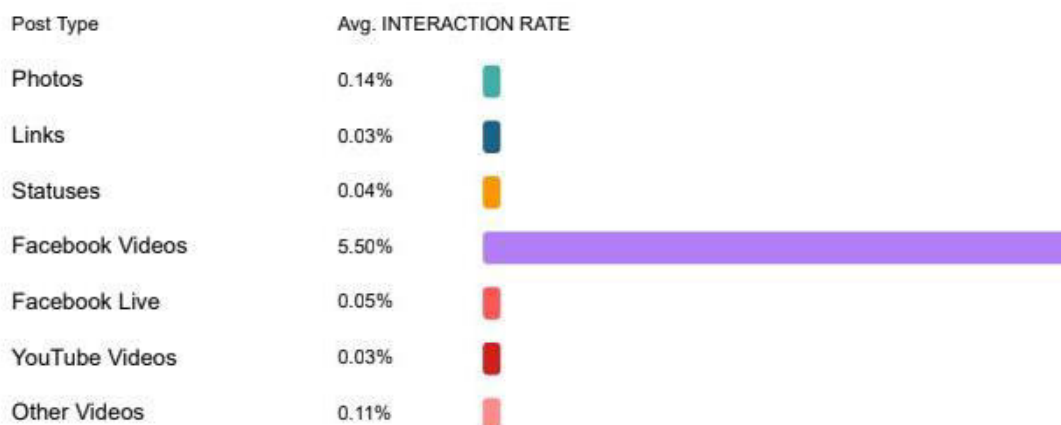
3.5 Comparative Analysis between Sectors

This case study is built on the list corresponding to the Instagram and pages European Cultural sectors analysed both in D1.4 and in the present report, namely: the Most Visited Museums, the National Libraries, the Archives, the Fashion GLAMs, the National Theatres, the Archeological sites.

The main goal of this analysis is to compare how institutes from different cultural sectors connect to their community of digital users in the temporal window from mid 2019 to mid 2021, observing how the Covid-19-related forced digitization impacted upon such relation.

Important disclaimer: as for the single sector's analysis, this comparative analysis is descriptive and all the observations and interpretations are correlations. We are aware of the possible bias that can emerge from the different weight and distribution of the compared samples.

Post count Instagram [figure 25 below] The graph shows the comparison between the six lines related to the total number of content posted per sector per month in the same time window, namely July 2019 - July 2021. The highest number of digital contents has been reached by the Museum sector (which seems to be the most active and productive, even tho they deal with pictures rather than texts and, in a particular situation like the lockdown, it's easier to think that a digital user is more attracted from visual (nice) objects rather than a digitsed manuscripts. It is also known that Theaters had serious issues of copyright to deal with because they manage contemporary productions and consequently the release of digital content is more difficult) followed by the Libraries sector and the National Theatres.

For all of the sectors, the highest increase in the number of posts produced corresponds to Spring 2020, namely to the first lockdown period, and the second peak corresponds to the second lockdown period, namely around October 2020, with a delay by the Archeological sector.

Name	Total	%	
Fashion GLAMs	7.2K	11.34%	
Most Visited Mus...	23.2K	36.72%	
EU Libraries	12.3K	19.40%	
EU Archives	7.2K	11.42%	
EU Archeologic S...	2.6K	4.05%	
EU National Thea...	10.8K	17.07%	

Interaction rate Instagram [figure 26 below] in this comparison, the only observation that can help to understand a general trend is that all of them decreased along the last two years.

Post count Facebook [figure 27 below] The graph line shows the comparison between the six lines related to the total number of content posted per sector per month in the same time window, namely July 2019 - July 2021. The highest number of digital contents has been reached by the European Libraries sector, which proves to be the most active and productive on Facebook, followed by Archives and Theatres. Here again, for all of the sectors, the highest increase in the number of posts produced corresponds to Spring 2020, namely to the first lockdown period,

and the second peak corresponds to the second lockdown period, namely around October 2020.
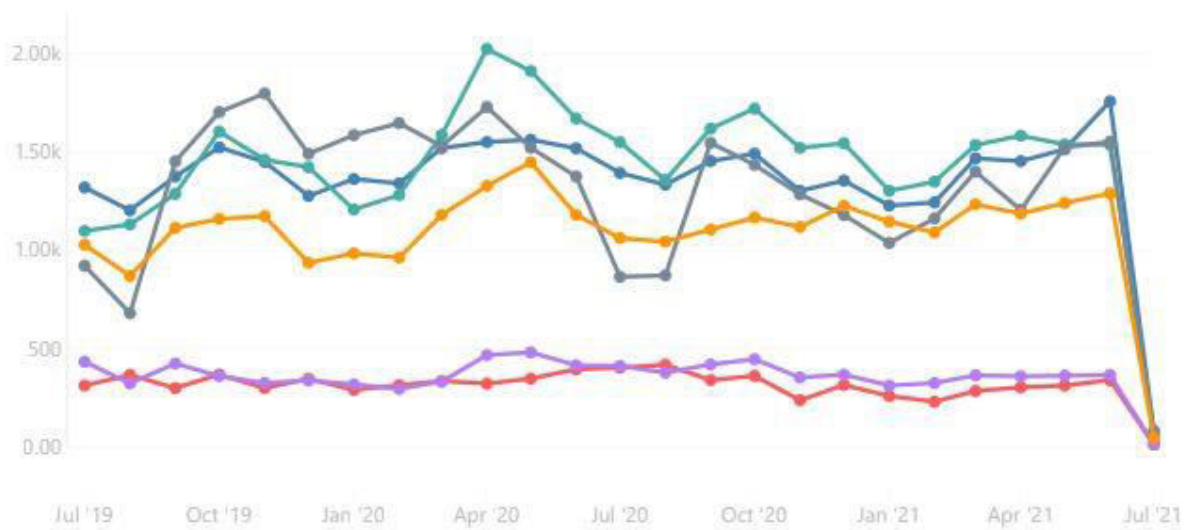
ALL POSTS > COMBINED TOTAL POSTS > 07/01/19 - 07/01/21

	TOTAL POSTS	AVG. MONTHLY POSTS
	146.8K	978

Name	Total	%	
Eu Archives	34.1K	23.23%	
EU Libraries	35.9K	24.47%	
EU Archeologica...	7.8K	5.34%	
Fashion GLAMs	9.0K	6.14%	
EU National Thea...	32.6K	22.20%	
Most Visited Mus...	27.3K	18.62%	

Interaction rate Facebook [figure 28 below] also for Facebook, the only observation that can help us to understand a general trend is that all of them decreased along the last two years with a general slight increase before the second lockdown period.

43
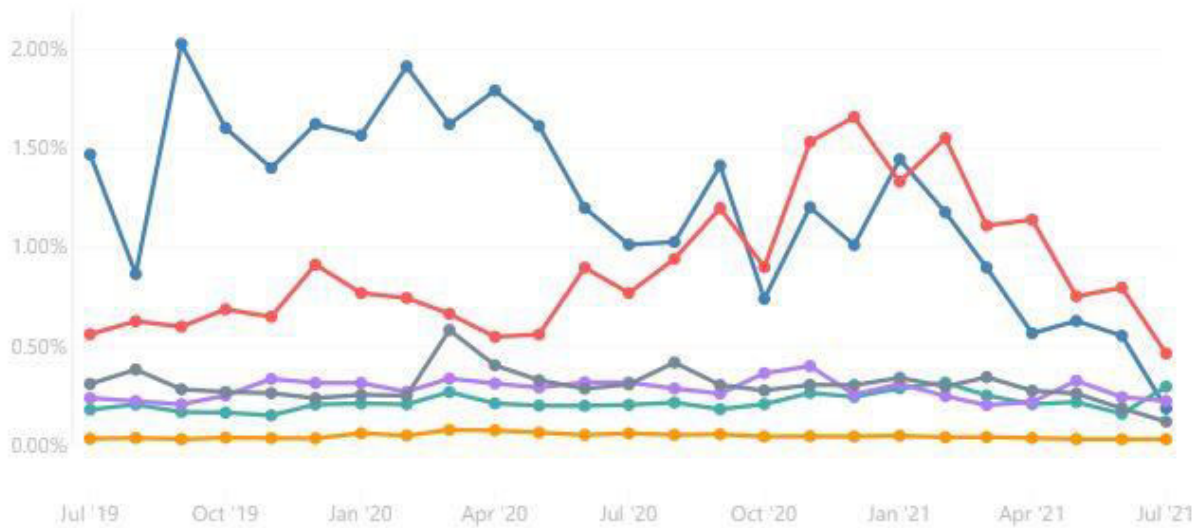
ALL POSTS > COMBINED AVG. INTERACTION RATE > 07/01/19 - 07/01/21

MONTHLY INTERACTION RATE	AVG. MONTHLY INTERACTIONS	TOTAL POSTS
0.49%	459.7K	146.8K

Name	Avg. INTERACTION RATE	
Eu Archives	1.22%	
EU Libraries	0.22%	
EU Archeologica...	0.89%	
Fashion GLAMs	0.28%	
EU National Thea...	0.30%	
Most Visited Mus...	0.05%	

3.6 Update on WLT Web Content Datasets

This section provides an update on the datasets gathered by WLT, previously reported in D1.4.

Throughout the reporting period data gathering has been ongoing. The WLT repository has been continuously mirroring and enriching the website content of

major News and Web Sources, as well as content from custom inDICEs-specified sources covering the cultural and creative sector. This data collection also includes social media content of specified Twitter, Facebook and Youtube outlets, made available for aggregated analysis and reports.

The Social Media Sources in particular were significantly extended, adding sources from the case studies in Section 3, and a new source filtering feature was developed to improve the search results.

3.6.1 Content Sources

News and Web Sources. Between 1 September 2021 and 30 November 2021, a total of 6.7 million news articles were gathered and processed across all six languages (en, de, fr, es, nl, it) and are now available via the inDICEs dashboard and the APIs. From 1 December 2021 to 14 March 2022 further 7 million news documents were collected. In summary, roughly 6.5 million articles are continuously gathered each quarter.

Over the full period the Web Corpus has grown by more than two million new documents, from both the inDICEs-specific web sources of the cultural and creative sector, as well as generic web sources provided by partner WLT which cover the public debate outside the News domain.

Social Media Sources. Since the last reporting, in addition to Twitter, we have broadened the data collection by including Facebook pages and YouTube channels in our social media gathering activity. The social channels added came from the cultural institutions that were collected by partners in WP1 as part of compiling the case study lists. The added Facebook pages cover many of the National Libraries and Archives that were considered in the analysis in D.1.4, as well as accounts of the National Theatres, Archeological Sites and Fashion GLAMs evaluated in Section 3.2-3.4.

Since we are only processing content from open Facebook "Pages", i.e. organization pages, and not from personal profiles, the listed sets are not fully covered, excluding organizations using a personal profile. Additionally, the Twitter accounts from which we are aggregating content were extended and now also include accounts from the selected lists of National Libraries, Archives, National Theatres, Archeological Sites and Fashion GLAMs, now totaling 211 accounts. From YouTube, we hold a collection of approximately 90 channels taken from the different case study lists. The selection of Dutch Twitter Sources has been supplemented by Facebook pages and YouTube channels of the respective Dutch Websites where they could be identified.

In total 27.5 million social media posts were captured, annotated and stored in the WLT repository between 1 September 2021 and 31 March 2022, the large majority of which stem from Twitter, with a comparably smaller yield of 27,680 Facebook postings and about 850 YouTube videos (almost exclusively from the Dutch Sample). This reflects both the specific selection of channels according to inDICEs requirements and the API characteristics and restrictions of the included social media platforms.
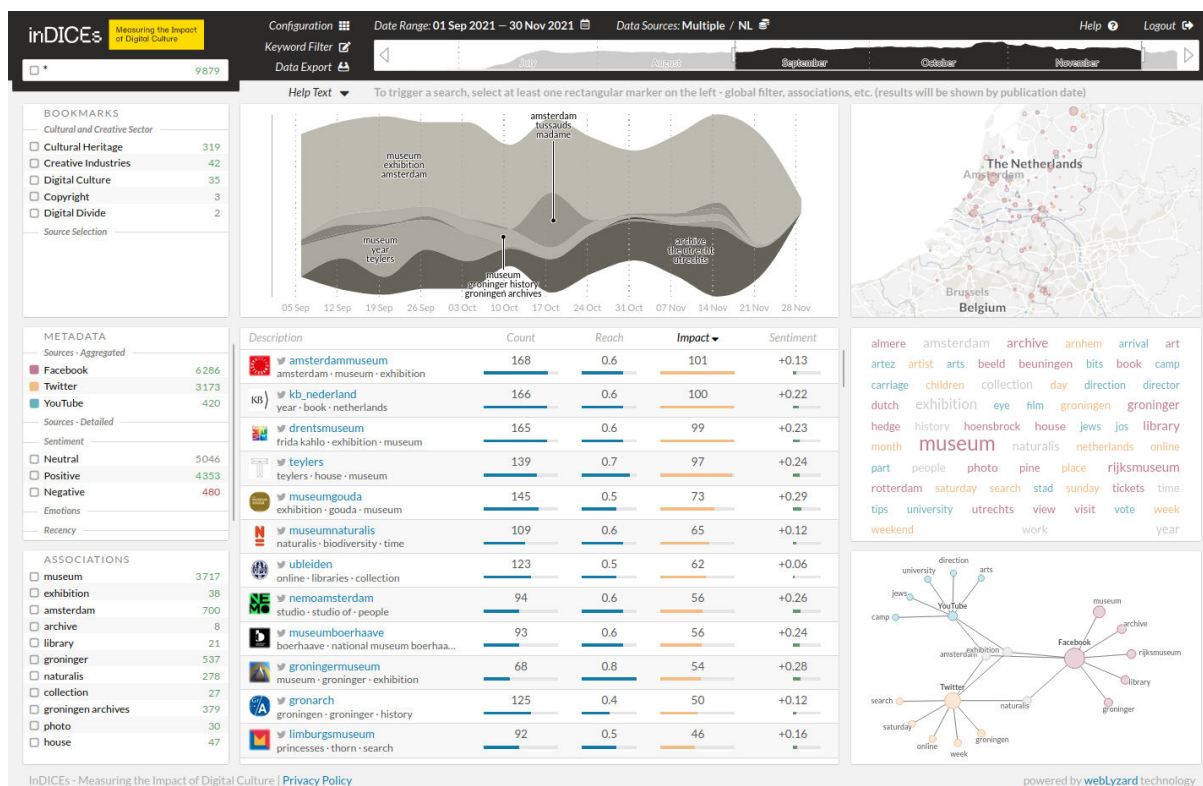
Figure 29: View of the Dutch Social Media Sources and their content between Sep and Nov 2021

3.6.2 Revised Topic Selection and Source Filtering

This section summarises work aimed at improving the searchability of the gathered data. Since the last reporting, the previously existing three main topic filters (*Creative Industries, Cultural Heritage, Digital Culture*) have been completely reworked and the new topic *Digital Divide* was introduced.

- *Creative Industries* now consists of 27 regular expressions that capture key phrases related to the different creative sectors, i.e. performing and visual arts, music and cinema.

- *Cultural Heritage* has been extended to 26 regular expressions associated with cultural assets, cultural consumption and cultural practices.

- *Digital Culture* now comprises a list of 13 regular expressions related to virtual experiences, digital assets and digitization.

● *Digital Divide* (created from scratch) captures any mentions of different forms of digital divide, including gender divide, social divide and digital access divide.

Edit Bookmark: Cultural Heritage >
List of Phrases

#	REGEXP	COUNT	
1	(antique|ancient|historical|historic|antiquarian) (building|art|piece|exhibit|construction	107	
2	decoloniz(e|ing) a? museums?	2	
3	cultural (appropriation|diversity)	120	
▼	artistic (expressions?|values?|heritage)	40	⚙ ✕
5	heritage assets?	7	
6	cultural legacy	10	
7	monuments?	791	
8	(works? of art | artworks?)	1263	
9	cultural (centers?|centres?)	141	
10	cultural propert(y|ies)	18	
11	museology	1	
12	(art|culture|cultural) (conservation|preservation)	10	
13	heirloom	93	
14	cultural (heritage)? objects?	12	
15	archeology	18	
16	museum acquisition	0	
17	fashion heritage	0	
18	booklover|architecture lover|lover of architecture	1	
19	art(istic)? patronage	1	
20	cultur(e|al) (elite|hierarch(y|ies)|circle|academy)	12	
21	cultur(e|al) (fund(ing|s)? | grants?)	4	
22	(culture|art) conservation	2	
23	cultural and creative industr(y|ies)	14	
24	cultural (producer|consumer|prosumer)	1	

Apply   Cancel

Figure 30: Excerpt of the list of regular expressions used to define the *Cultural Heritage* topic.

In addition, new bookmarks are now available that comprise pre-defined sets of sources related to cultural topics. They are available under the Bookmarks section *Source Selection* and currently consist of *Fashion* (the *European Fashion GLAMs*), *Contemporary Art and Art Market, Movie Festivals* and *Dutch GLAMs,* the latter being the result of the pilot conducted where a list of Dutch GLAMs were extracted from Wikidata and reviewed by project partners. While the complete list consisted

of 2194 Wikidata references with associated metadata – 1527 of which contained a link to their website – a final selected set of 77 institutions was added to the inDICEs Knowledge Graph. This set of entities is now fully integrated into the entity recognition and enrichment process, surfacing associated mentions of the respective institutions. With the new bookmark the content posted and distributed by the associated Dutch sources can be quickly and easily tracked across all media channels.

In the ongoing effort to establish the final set of bookmarks, which will be available as predefined search configurations in the Dashboard Light, project partners were further asked to communicate their preferred top keywords or phrases from the new bookmark definitions. This feedback will be used to further narrow down the final topics. WP2 contributed by specifying key terms for a "legal"/"copyright" related topic, which is also planned to be included. Simultaneously, the topic "Archeology" was explored in a first small work group using the inDICEs dashboard, and established to be of interest.

To generally allow for better filtering of the cultural sources we have introduced a new way to sub-categorize sources within the News and Web Sample. Categories are assigned in many-to-many fashion, which means that multiple categories may be assigned to a single source. For a first version we agreed on the categories in Figure 31.

skbtag:gallery	skbtag:film
skbtag:library	skbtag:theater
skbtag:archive	skbtag:music
skbtag:museum	skbtag:site
skbtag:art	

Figure 31: Newly introduced categories for cultural sources.

The intention for this selection was to a) keep the individual categories general, clear and re-usable, but b) have them aligned with the inDICEs case study lists. The categories – which we will also refer to as "tags" – are aligned with the WLT knowledge graph model and associated with their individual LOD URI in the WLT namespace to allow referrability and reusability.

A first simple algorithm was applied to automatically assign tags to the individual website sources collected by inDICEs. For this, a set of regular expressions was composed for each category that captures different spellings and language varieties, as well as singular and plural forms. The sets were then cross-checked for matches in the website domains and associated source names. For the sources collected by inDICEs partners the assigned cultural sectors resp. cultural regimes of production, as formulated by WP1, were also considered. For the sources extracted from Wikidata for the Dutch pilot case, not only the URLs and labels (`rdfs:label`), but also the descriptions (`schema:description`) and Wikidata classes, provided through the `instance of` (`wdt:P31`) property, were considered. While the first eight categories refer to the contents that a web source publishes or represents, the last category *skbtag:site* is assigned whenever an associated geographic location is known for a source, for example the physical location of a museum or a corporation's headquarter location.

With this strategy we were able to assign at least one tag to 1,016 of the total 1,632 CSS sources, and at least one to 1,514 of the 1,725 Dutch GLAM sources, the second being easier to assign due to the additional metadata extracted from Wikidata.
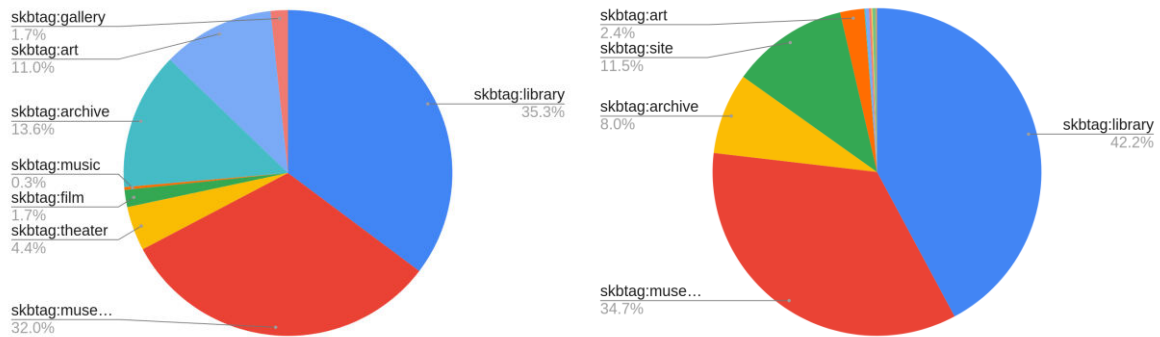
Figure 32: Comparison of the automatic tag allocations (multi-assignment possible) for CSS sources (left) and Dutch GLAM sources (right).

To further assign categories to website sources that were still missing a tag, partner PIN SCRL assisted, and manually revised and curated the assigned tags. With this work an additional 331 CCS websites could be assigned a tag from the established categories in Figure 31 and can be filtered effectively. Furthermore new suggestions for additional categories were proposed, wherever existing ones were not sufficient. Those categories will be further discussed and refined in the coming weeks.

From a user perspective, four new sub-selections are available in the inDICEs dashboard for the MISC/Web Data Source menu that allow multi-filtering for the respectively tagged sources only. The first four categories, consisting of websites of galleries, libraries, archives and museums, were combined into a single *GLAM* sample. Similarly *Film* and *Theatre* were combined into a single selection.

Combining those new source categories together with topic bookmarks allows for very flexible search queries, where the visualized documents can be restricted to a certain sub-category of interest and further narrowed down with context filters.

Figure 33: *Story View* cluster representation, showing content with active filters of all new Data

Source Samples (*GLAM*, *Art*, *Film/Theatre*, *Music*)

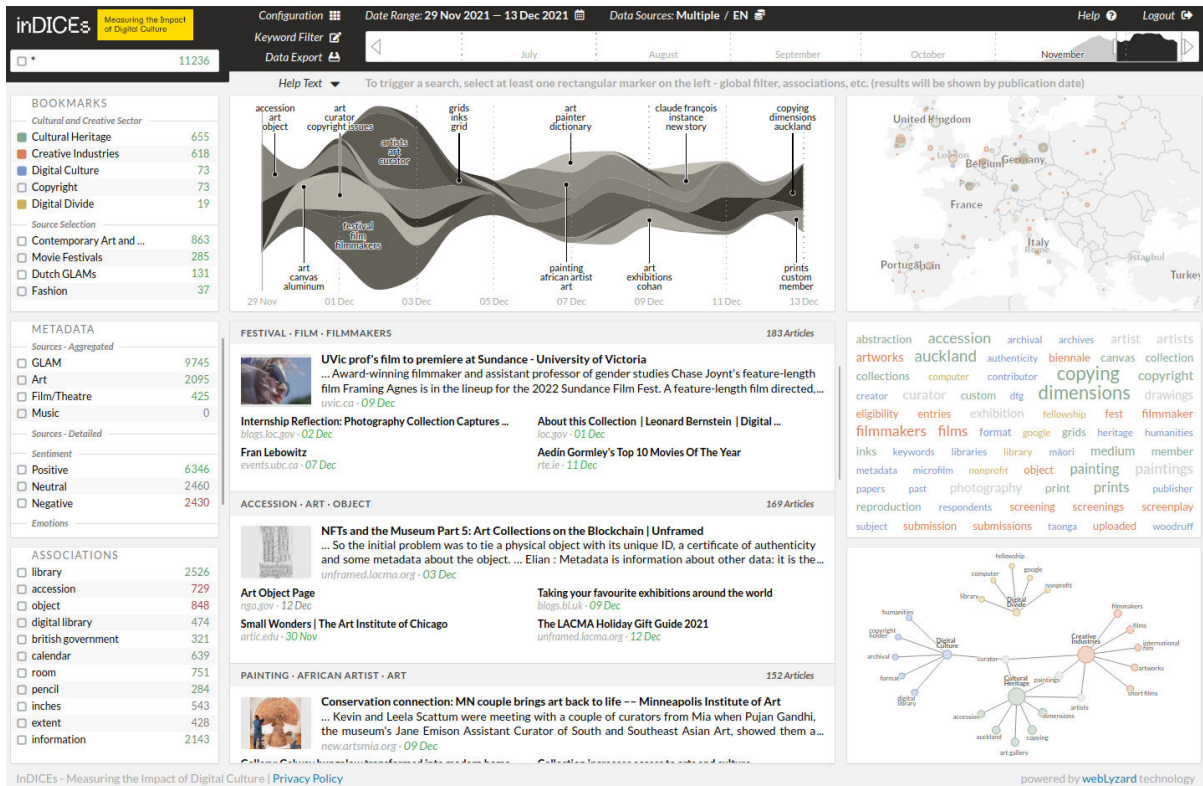# 4. Preliminary observations

The aim of this paragraph is to give a partial overview, according to the available data gathered and analysed until Month 24 and to the data gathered and analysed that have been presented in the D1.4, aimed at detecting the trends regarding the levels of digital cultural participation of the users of the institutions of different sectors', regarding the sectors that we have seen in the "Data gathered to Month 24" section.

For what concerns the Fashion GLAMs, in general, we can observe how the pandemic influenced the interaction of their social network users, and how strongly the sector's production of contents and relation with its digital communities passes through the live events. Indeed, the Instagram interaction rate strongly decreased in the last year, while before the first lockdown period we can observe four peaks of interactions; again, regarding the rate of the views, we can observe how the highest peak corresponds to the period in which the world of Fashion restarted to host and produce live events. This can be confirmed also by the typology of tool preferred: Videos and IGTV are the two main typologies of contents visualized, but videos, which are more likely to be pre-recorded and with edited contents, gain more visualizations than IGTVs, which is a tool that is more likely to be live-streamed but couldn't be exploited for Covid-19 policies reasons. Anyway, this element indicates a great effort toward digitization of archival contents. Even on Facebook, the event sector and consequently the live-streaming channel is quite invisible and this is the most evident effect of the social distancing policies. We must observe that the effort implied to digitize contents, in order to keep alive the relationship between Fashion GLAMs and their users, has been high and fruitful: social networks of Fashion Institutes gained followers, attention and an increasingly high interaction as soon as the physical presence has been allowed again.

Also in the analysis of the data regarding the European National Theaters, we must consider the symbiotic and necessary relationship with live performances. The number of new followers gained on Instagram over the last two years increased a lot, in particular, we can observe the highest growth in two time windows tat corresponding to the first lockdown period and winter 2020-21, but that can also be correlated with Christmas - New Year time. On Instagram, Photos and Albums are the content type generating most of the interactions, even if stable images reduce the sphere of movement as sound and actor performance that a Video can catch and reproduce. Video is the most viewed channel of content sharing, showing that European National Theaters didn't pay much attention to their users' preferences. IGTV, due to the policy reasons listed above, is not even considered both by the National Theaters pages and (consequently) by their users.

On Facebook, the interaction rate percentage is higher considering Videos and Live streaming, with various peaks all along 2020 (posted videos have been of course the vast majority, with various very high peaks corresponding mainly with the first lockdown, since most of the Live Videos couldn't have been produced for lacking of live performances.) Users' attention and interest has been kept alive probably by digitized archival materials, showing an effort toward content digitization.

The case study built on the list of the Facebook and Instagram accounts of the European Archeological Sites is peculiar in terms of the small size of the sample and the non-homogeneous distribution of the countries that host Archeological Sites having social network pages. As far as we could observe by a direct observation of the Instagram and Facebook pages of this list, a real bottom-up digitalisation occurred, a spontaneous process of co-creation of content with which cultural institutions could not, and may still cannot keep up. In general, most of the contents are directly created by users at the moment they are visiting the site, by posting pictures, videos, live videos or by geolocalizing themselves in the Site. Moreover, we can observe that before the first wave of pandemic Instagram pages of European Archeological Sites didn't even consider Video as a tool for sharing contents: they started to produce video content around the end of February 2020.

Focusing on the comparison between different Heritage Sectors analyzed both in D1.4 and in the present deliverable, the graphs show the comparison between the six lines related to different activities per sector per month in the same time window, namely July 2019 - July 2021. In general, we can affirm that it is not easy to obtain a meaningful comparison of the graph lines of the interaction rate given the significant differences between the samples. The only observations that can help us to understand some general trends are the following:

- the pandemic and the lockdown periods in particular led CHIs to an overproduction of digital content. The highest number of digital content has been reached by the European Libraries sector, which proves to be the most active and productive on Facebook, followed by Archives and Theatres. Here again, for all of the sectors, the highest increase in the number of posts produced corresponds to Spring 2020, namely to the first lockdown period, and the second peak corresponds to the second lockdown period, namely around October 2020.

- in the last two years, all the sectors gained high percentages of new followers, mainly on Instagram

- the interaction rate of all the sectors decreased along the last two years, and, as we have already noticed in D1.4, if the number of followers grows and the interaction rate drops, it is plausible to conclude that the new followers are inactive and only the hard core of already registered and active users continues to really interact. In our samples, despite a constant follower growth, the interaction rate decreases, revealing that the real uptake of the Culture 3.0 production regime still seems to be in its early phase. It is necessary for CHIs to better enable people to actively participate in meaningful sense-making processes, to exploit the possibilities that the digital platforms can offer in terms of co-creation processes, digital community empowerment, development of new soft skills and shared knowledge resources.

5. Usefulness and Limits

The information of the data analysis that is reported in the present document is likely to be useful for a wide range of potential users and stakeholders in cultural and creative sectors, thanks to the data accessibility as described in deliverable D1.2, through the inDICEs Open Observatory, in conjunction with the embedded widgets of the Visual Analytics Dashboard.

For what concerns the limits of these analyses, it is important to underline that our data are aggregated, and consequently they need to be eventually unpacked and parsed in greater detail according to the specific interests of the inDICEs partners and stakeholders.

6. Early results on users behaviour in content co-creation

In this section, we will briefly explain the inDICEs analysis regarding the study of user behaviour. In particular, we will show the cases of Wikipedia and TikTok, which, as we will see, are opposite for different reasons of both moderation and type of content.

The aim of the present analysis, conducted by FBK and reported in the present deliverable only partially for reasons due to peer-review ongoing process, is to investigate, on one hand, one of the most significant examples of participation and collaboration that we considered with special attention, namely Wikipedia, and its virtual knowledge communities; on the other hand, we aim at exploring the hierarchical organisation of the TikTok ecosystem, in order to understand how the creative content re-mix, re-use and production happens in one of the most relevant digital social network of the current digital sphere.

These analyses are embedded in the investigation and enrichment of one of the most relevant inDICEs theoretical frameworks, namely the theory of the 3.0 Culture, which regards the contemporary strategies and dynamics of cultural production and reproduction in the web 2.0. These analyses can contribute to orient the Policy Guidelines for CHIs and the European Policy Makers to support them in understanding the contemporary evolution in terms of digital users behaviour, in relation with cultural and creative creation, re-production and with the dynamics of virtual community development - processes that may pave the way for the evolution of Digital Heritage Communities.

6.1 Wikipedia

The application of collective intelligence theories on Wikipedia is linked not only to the participatory culture that moves users to collaboratively publish and edit content but also to the use of open-source standards, such as the "wikis". Founded in

2001, the publicly editable online encyclopaedia Wikipedia is likely to be the most exhaustive and up-to-date repository of knowledge in the world. More than an online encyclopaedia, it is a platform based on the wiki approach that encourages volunteers to edit pages on the website at their own discretion, without any central authority and without the need for special software. In the past fifteen years, hundreds of thousands of volunteers around the world have contributed their time and expertise to the project and as a result Wikipedia is currently home to over 45 million articles. While Wikipedia started off in English, it is presently available in almost 300 languages. The Wikimedia Foundation (WMF), the organisation running the sites, reports that every hour 15,000 edits are made and that over 10,000 articles are added each day. Since 2007, "the free encyclopaedia that anyone can edit," as it describes itself, has consistently ranked in the top ten most visited websites of the world, despite being the only non-profit and primarily volunteer-based organisation of this list.

With the introduction of Wiki software the asymmetry of the content production scheme on the Web has been reversed, moving it from the client to the server. As a consequence, a page that before could be read by everybody but modified only by the owner, now becomes a page modifiable by anybody through the browser, inserting the appropriate markup that will be interpreted and translated by the server side software that will generate the corresponding HTML page. In Wikipedia on average there are ten edits per minute. The total number of authors per page is a difficult value to measure because a single author may have multiple accounts and Wikipedia allows anonymous editing (IP address is tracked but this is not a value with characteristics that can be associated with a specific user). Wikistats measurements are based on active Wikipedians defined as those users who have made at least five edits in the last month. But Wikipedia is not simply an online encyclopaedia that allows everyone to contribute, the site is actually the most visible artefact of an active and emerging community. Wikipedia is first and foremost a community and the encyclopaedia is the image and result of the discussions of this community is the realisation of the collaborative culture in search

of a universal encyclopaedia. As a project of online collaboration, it represents an excellent example of social production of knowledge (reflecting the interests of the population of users) and the study of the agents that produce the content of Wikipedia is, therefore, a topic of discussion of great interest for both humanities and the social sciences. In this regard, several studies have been carried out over the years, following different approaches and methodologies, and even conflicting results have been obtained. Within these studies, the most dazzling result was a comparative analysis in terms of accuracy of scientific pages between Encyclopaedia Britannica and Wikipedia. The study conducted by a Nature[7] team consisted of asking academics to analyse 50 pairs of articles extracted from Wikipedia and the Encyclopaedia Britannica website, without knowing which of the two encyclopaedias each article came from. Each respondent highlighted a list of errors for each article, a total of 123 for Britannica and 162 for Wikipedia. The scholars' comments on the errors were then associated with numerical values by Nature staff. Based on these results, the study concludes that the information contained in Wikipedia can be considered as reliable as that of the Britannica and that errors in Wikipedia would be the exception and not the rule. This result in itself does not diminish the reliability of the Britannica, but highlights the high level of accuracy of Wikipedia, a feature that was not suspected to be at such high levels before this comparison. The answer to why Wikipedia is so reliable does not seem to lie in the security provided by the existence of guidelines, correction rules, or patrolling systems that actively address errors, but rather in the approach of Wikipedians to the system, who contribute articles to the encyclopaedia out of a spirit of collaboration and a desire to be part of a community whose goal is the creation of common knowledge. In conclusion, it is possible to highlight how Wikipedia is the perfect example of the effectiveness of open platforms in creating cultural content of the highest level but at the same time how the wiki ecosystem still presents the typical digital platform traits of asymmetry in content co-creation.

---

[7] see Giles, J. Internet encyclopaedias go head to head. Nature 438, 900–901 (2005). https://doi.org/10.1038/438900a

Figure A: measures of probability in a non-representative subsample of the Italian Wikipedia dataset (2001-2020)
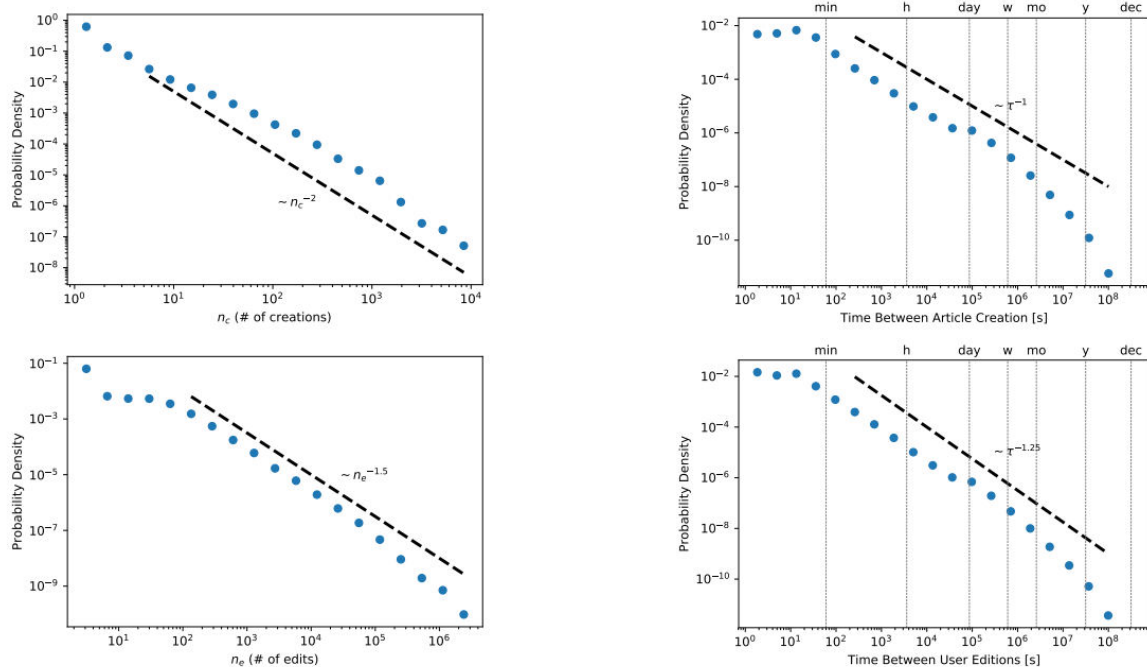
The statistics of content creation in Wikipedia indicate a very heterogeneous user behavior. This results from the probability distribution of two quantities at single user level (i.e., the number of article creations and editing). These distributions are shown in Fig. A (top row), and both of them are fat-tailed. The power-law behaviour —a more restrictive case of fat-tailed distribution— seems to approximately hold only in the region of large values of the quantities considered, while for low values it seems to break down. We might run statistical tests to verify how much compatible are these distributions (or their tails) with power-law behaviour, but probably the message should not be whether or not we have a power-law, but that these distributions are telling us that the average human behaviour in the Wikipedia ecosystem is not well defined in the sense that the deviation from the average is huge (fat-tailedness). Regarding the edition dynamics, the message is the same: fat-

tailed distributions, see Fig. A (bottom row). When compared to the creation of articles, several things can be said, all of them pretty straightforward though: the probability function of the number of editions is asymptotically equivalent to an exponential with (negative) exponent -1.5, while the number of creations is asymptotically equivalent to an exponential with negative exponent -2, what makes the former decaying more slowly than the latter. This may be interpreted as creations being more difficult than editions. Note also that not only the decays are different but also the range of variables: creations go up to 10^4 while editions go up to 10^6, which simply is an alternative way of seeing that the engagement of users is higher when they participate in something that is already created.

Figure B: unique use in a non-representative subsample of the Italian Wikipedia dataset (2001-2020)
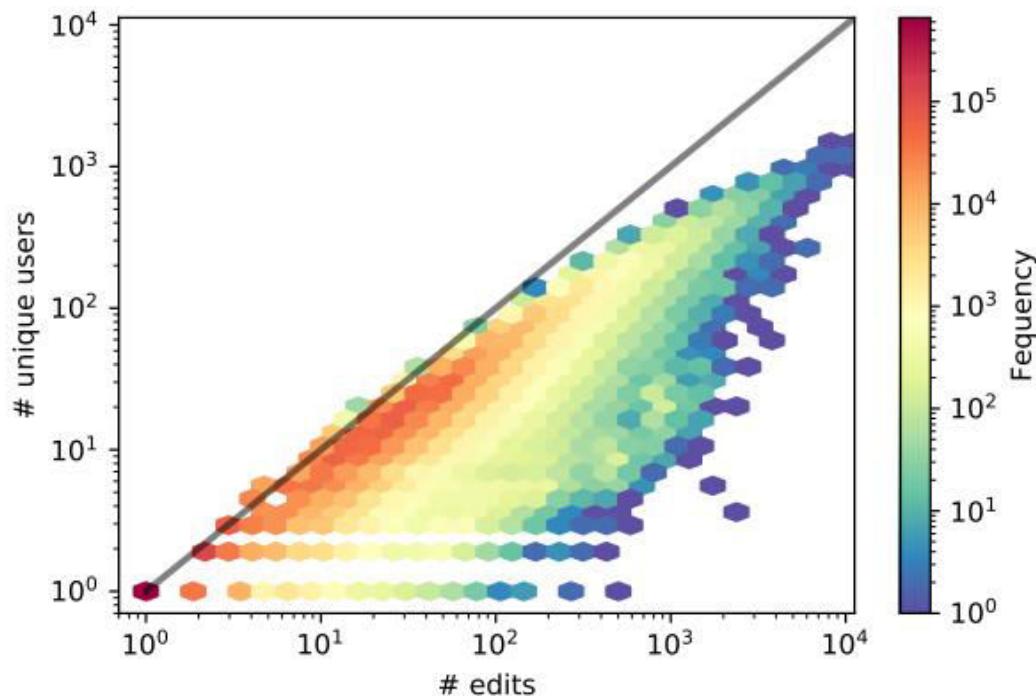
We can also address the collaborative editing of articles. Quantitatively, we can look at the number of unique users as a function of the number of edits an article has. We show this in Fig. A, and we see that for low and intermediate values of the

number of edits, the most probable outcome is that we observe as many unique users as the number of edits. Therefore, the collaborative effort is almost maximum. This tendency decreases a bit for larger values of the number of edits, meaning that for frequently edited articles there is a higher probability to observe the repetition of users in the editing. However, we see that it is highly unlikely to observe frequently edited articles by very few users. This result manifests the crowd-sourced nature of Wikipedia when it comes to review (and probably improve) the quality or accuracy of articles.

6.2 TikTok

(Disclaimer: it is a merely qualitative approach which give some insight into the phenomena under study.)

Within the current media and digital landscape, TikTok has surely gained notoriety and global relevance for producing and sharing massively creative contents by crossing music, re-used images, and new short-video contents. In September 2016 the company released the Chinese version, called Douyin, while TikTok was launched in 2017. Finally, in 2018, TikTok expanded its market and user base, by merging with the already popular social media app Musical.ly. With about 1 billion monthly active users worldwide, TikTok owes its popularity to teenagers and young adults, who, according to recent statistics, account for over 60% of the user base2.While TikTok became popular to the public for dance and lip-sync videos, existing research has demonstrated that theplatform also represents a hub for youth to discuss political and socially relevant issues. This has become apparent in the context of the Covid-19 emergency, during which users used TikTok to spread information and awareness on how to prevent the virus. Existing research has offered different explanations to account for the popularity of the platform. Some argue that, unlike other platforms, TikTok emphasises content production over consumption6. Accordingly, its technical infrastructure is designed to foster users' creativity and self-expression, appealing especially to young people: in this context,

it has been argued that the editing features afforded by TikTok reproduces the elaborate cinema professional post-production in a 'virtual playground setting. In this sense, TikTok provides users with a rich toolbox that integrates sophisticated technology to offer different possibilities to creatively engage with content creation, including "in-camera speed controls, image-tracking composites,collaborative split-screens". From another perspective, a growing body of studies demonstrate how the platform is designed to grab and maintain the attention of the user, leveraging a carefully customized algorithmic-driven mechanism of content curation. Once logged in, TikTok users are immediately presented with a video and the decision to watch (and interact with) it or scroll down will affect the quality of the future videos shown in this section, which is not casually called "ForYou Page" (or FYP). FYP is sensitive to both users' engagement and a variety of collected personal data, such as the geolocalization. The result is a feed which is continuously adjusted and refined to keep users interested and engaged. Although the functioning of TikTok algorithm remains mostly obscure, it appears to be more centred on aggressive promotion of content rather than on building a social network infrastructure. From these studies it also emerges that TikTok fosters the connectivity of users and content: users are encouraged to interact with existing content not only as audience, i.e. by commenting, liking and sharing, but also through the creative practices of re appropriation and repurposing, a dynamic which appears to play a crucial role for the emergence and the diffusion of viral trends. In this context, the 'templatability' of native digital objects like filters, stickers, and sounds contribute to the idea that TikTokstructure is inherently memetic. TikTok has caused a significant paradigm shift in the influencer industry, giving prominence to relatable and entertaining content. Valuing performance over "Instagramesque physical appearance", TikTok seems to have crafted a new conceptualisation of influencers sidestepping aesthetically pleasing content to focus on more "personalised disclosures and storytelling", not ultimately affecting the perception of the accessibility perceived by users towards influencers. Along This line, it has been argued that younger users look at the platform as an opportunity to showcase their

creativity and their talent, receive feedback and acquire visibility. Visibility labor on TikTok relies on a number of interactive and algorithmic practices, aimed at increasing the chance of going viral. Another widespread practice observed consists in tagging the video with the hashtag "foryou" or other popular hashtags, in the hope to push it into the trending section. Finally, users believe that even posting videos at certain times may increase the chances of reaching higher visibility. Our interest in the study of TikTok concerns its double role in creating communities of practice around socio-cultural trends (such as the aggregation of users around the Black Lives Matters movement or K-POP) and its peculiar characterization from the point of view of memetic affordances, that are reified in the reuse of cultural objects as short musical extracts (think in this case at the phenomenon of choreography challenges that have become famous worldwide). In the light of the Culture 3.0 paradigm, TikTok presents itself as the ideal platform both to understand how users interact with each other in the creation of new cultural content and to understand how affordances and algorithmic drivers create and structure sub-genres that are typical of the memetic and recycling sub-culture which characterizes the platform society and mediatic convergence.

Figure C: Quantifying popularity. On the left, a normalized histogram of the number of comments posted in the videos. On the right, In both figures, the data have been logarithmically binned for a cleaner representation.
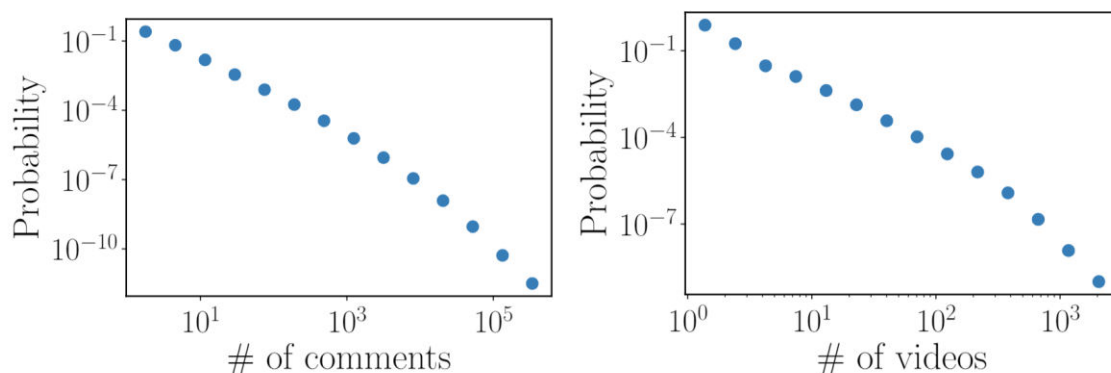
We compute several metrics to assess how information and popularity is distributed across users and posts in our dataset. In the left panel of Figure C we show the normalised histogram of the number of comments in the videos of our dataset. This gives the probability that a randomly selected video has a certain number of comments. Put otherwise, the histogram gives the fraction of videos with a certain number of comments. We observe that this quantity decays slowly, much slower than exponential decay, and spans several orders of magnitude. If we take the number of comments as a first proxy for the popularity of a video, this result evinces the huge variability underlying the content creation in TikTok. We will discuss later on more implications of such a broad distribution. To further explore the hierarchical organisation of the TikTok ecosystem, we also compute the normalised histogram of the number of videos posted by individual users, see the right panel of Figure C. We observe that, again, this quantity is distributed across several orders of magnitude, with many users posting few videos but also with a non-negligible amount of users that post many. However, these two behaviours do not represent the only two possible modes of creation in TikTok, but there exist a continuous, self-organised spectrum of behaviours interpolating these two extremes. The presence of such heavy-tailed distributions indicates the difficulty to quantify a representative, average behaviour behind the social platform we are studying. However, at odds with rapidly-decaying distributions, the peculiarity of heavy-tailed distributions is that the moments might depend on the sample size and diverge when the number of measurements is large enough. Therefore, it is difficult to extract well-defined statistics of the system under study due to this dependence on the sample size, because if we use a larger pool of videos or users, the moments of the distribution can change.