TRANSMIXR

IGNITE THE IMMERSIVE MEDIA SECTOR BY ENABLING NEW NARRATIVE VISIONS

D2.2: Media content metrics and visual analytics dashboard

Lead author/organisation: Arno Scharl, webLyzard technology (WLT)

Co-authors/organisations: K. Böhm, D. Fischl, M. Göbel (WLT), M. Hospital, D. Ramond (AFP), M. Föls (Storypact)

Versioning (by)	Date	Version
A. Scharl | Created initial draft	25.01.24	0.1
A. Scharl | Refined document structure; introduction	08.02.24	0.2
K. Böhm, D. Fischl, M. Göbel | Details about platform and dashboard technologies in Sections 2 and 4	20.02.24	0.3

M. Hospital, D. Ramond | Pilot development, Section 3	21.02.24	0.4
M. Föls | Storypact screenshot, revision of Section 2	22.02.24	0.5
P. Cesar, A. Karakottas | Provision of feedback and suggestions	27.02.24	0.51
Arno Scharl | Incorporation of Reviewer Feedback	8.03.24	0.6
Arno Scharl | Final revisions and layout	14.03.24	1.0

Executive Summary

This deliverable outlines the content exploration and analysis capabilities of the TRANSMIXR dashboard, including the computation of content metrics and automated extraction of narratives from large document collections. The dashboard enables users to access and examine the outputs of content understanding, summarisation, and visualisation components developed in other WP2 tasks.

Central to the TRANSMIXR architecture, the dashboard is a key element of the experience creation workflow. Accessible to all project partners, it facilitates a comprehensive visual exploration and analysis of metadata-enriched content assets. The dashboard's ability to visually represent and navigate through content enhances the utility of pre-processed content from diverse sources, including the content of project partners such as AFP, RTV SLO or SPARK, as well as public resources from Web crawls and content feeds from third-party platforms such as YouTube.

The dashboard's utility extends across various aspects of the TRANSMIXR work packages, offering distinct versions (PRO and LITE) to cater to different user needs. It enables users to explore and select multilingual content assets, supports evolving use case scenarios such as the *Newsroom of the Future*, and provides market watch capabilities to track competitors and domain-specific news sources. Furthermore, the dashboard's embeddable widgets facilitate integration into third-party applications.

The dashboard and its content metrics - e.g. the keyness of a term, polarity of sentiment, or level of association with desired and undesired topics - are instrumental in pilot development, with user onboarding, requirements elicitation and customisation steps to tailor the tool to specific workflows. This deliverable outlines these steps and presents significant extensions and enhancements, including a multi-colour *Story Graph*, an improved entity inspection tooltip, a document highlighting feature, and upcoming prediction features for agenda setting and compiling editorial calendars.

This project has been funded by the European Union as part of the Horizon Europe Framework Program (HORIZON), under the grant agreement 101070109.

Table of Contents

This project has been funded by the European Union as part of the Horizon Europe Framework Program (HORIZON), under the grant agreement 101070109.

1. Introduction

Reporting on work conducted in T2.4, this deliverable introduces the first prototype of the *TRANSMIXR Visual Analytics Dashboard*, a visual frontend for exploring (i) the content ingested in T2.1, (ii) metrics to describe this content as well as its impact, and (iii) the results of the content understanding, summarisation and visualisation components developed in T2.2 and T2.3. Thereby, the dashboard enables professional users to identify relevant content, discover emerging stories and the opinion leaders driving these stories, and analyse observable patterns along multiple context dimensions (time, space, sentiment, etc.). To facilitate this analysis, interactive tooltips enable users to inspect metadata, trigger drill-down operations and request additional background information on a referenced entity.

The content metrics and the visual analytics dashboard presented in this deliverable are important elements of the interactive/immersive experience creation workflow shown in Figure 1. The TRANSMIXR dashboard is available to all project partners at transmixr.weblyzard.com. It provides visual means to access, navigate and analyse the metadata-enriched content assets from the "indexing and metadata extraction step". The content sources and their pre-processing have already been reported in Deliverable 2.1, including focused Web crawls and content feeds captured via the APIs of YouTube and partner AFP.

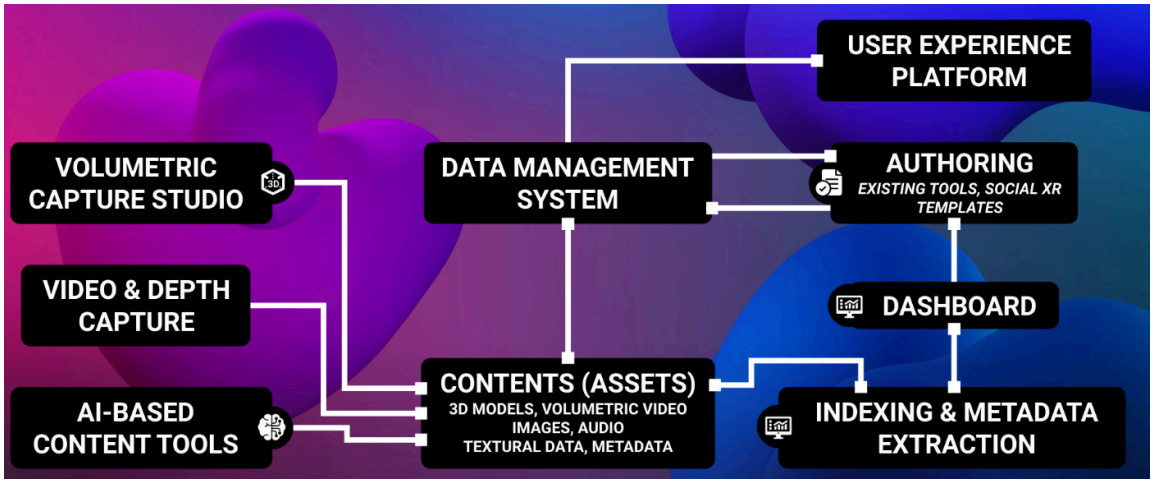

Figure 1. TRANSMIXR Dashboard embedded into the overall TRANSMIXR architecture.

The Visual Analytics Dashboard and its built-in visualisations serve multiple roles across several TRANSMIXR work packages (to cater to analysts and professional users as well as casual users or the public across these scenarios, the dashboard is available in a PRO and a LITE version):

This project has been funded by the European Union as part of the Horizon Europe Framework Program (HORIZON), under the grant agreement 101070109.

- **Explore and Select Content Assets** (for example, Sound & Vision Records,[1] Museum Collection Items,[2] AFP Articles,[3] etc.). After annotating/classifying assets by topic or region, users can select specific assets for editing in the *Creation Environment*. The dashboard will allow users to identify similarities within and across collections, automatically cross-link assets and visualise trends and patterns in evolving collections. → WP5: Pilots and Evaluation.

- **Support Evolving Use Case Scenarios** such as the *"Newsroom of the Future"*, a collaboration between the technical partners and AFP. The work in the first project phase has extended the dashboard as a tool for editors and journalists, for both internal assets and third-party content. This includes features to track emerging stories (Section 4), predictive features for agenda setting or defining a news organisation's editorial calendar. → WP5: Pilots and Evaluation.

- **Embeddable Widgets.** *Visualisation-as-a-Service* (VaaS) rendering engine for third-party applications, such as the Storypact Editor, partner Websites, or smartphone applications. → WP4: Media Creation.

- **Market Watch** to track and obtain a better understanding of competitors, new startups, domain-specific news sources, etc. The focused content ingestion in multiple languages is described in the "*Custom Content Feed on Immersive Technology*" section below. → WP5: Impact Metrics and Optimisation.

2. Dashboard Structure

Visual analytics methodologies merge the disciplines of data science, information retrieval and visual representation techniques to support content discovery, track emerging stories and generate insights for content producers and communication experts. The TRANSMIXR dashboard, developed by WLT, is an interactive tool that analyses perceptions and trends across stakeholder groups in online publications (media outlets, corporate entities, NGOs, the public sector, research labs, etc.).

The dashboard is a key component of TRANSMIXR's content creation and authoring workflow (see Figure 2). It enables users to identify relevant content, discover emerging stories, and analyse patterns across dimensions like time, space, and sentiment. It incorporates impact metrics to assess individual influence and integrates emotion detection algorithms for insights into stakeholder perceptions regarding organisational activities within specific contexts. Employing continually

[1] data.beeldengeluid.nl
[2] e.g., www.khm.at/en/visit/collections or www.britishmuseum.org/collection
[3] www.afp.com/en/search/site/articles

This project has been funded by the European Union as part of the Horizon Europe Framework Program (HORIZON), under the grant agreement 101070109.

evolving knowledge and advanced tools, including opinion mining and artificial intelligence algorithms, the dashboard provides real-time analyses of prevalent keywords, emotions, and topics, aligning with the objectives outlined in T2.4.

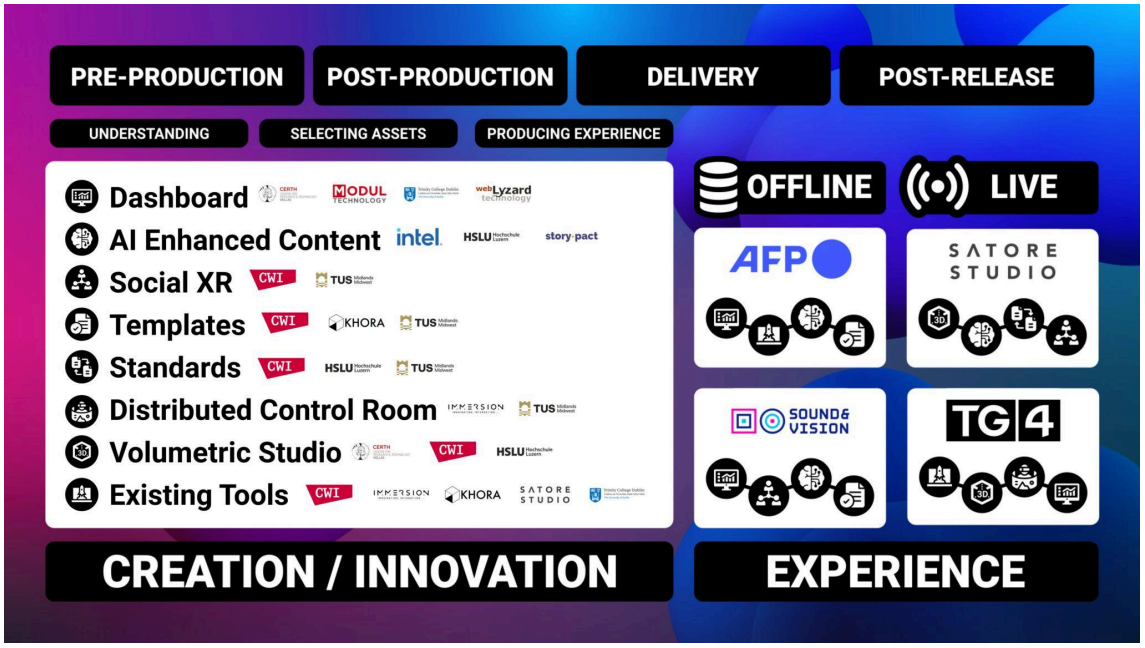

Figure 2: Content creation and authoring innovations embedded into the TRANSMIXR workflow

The dashboard's visualisations provide valuable insights, presenting search results in a format reminiscent of a classic search engine. Keywords are shown in a tag cloud for quick association views, and semantic relationships between terms are illustrated via a keyword graph and other visualisations. Moreover, the tool tracks topic trends over time and organises search results thematically and geographically. Users can customise searches using the dashboard's configuration menu, adjusting content filters, date ranges, and source types to suit their needs.

The dashboard lets users navigate the search space across different metadata dimensions, facilitating tailored exploration and a nuanced understanding of digital content — e.g., how communication campaigns engage target audiences, and how various topics are interpreted and framed by professionals and the general public.

To meet the project requirements elicited in WP1, the development process changed the general user experience, as described in the following paragraphs, and specific components such as the *Story Graph* (multi-colour capabilities, improved labels, etc.) outlined in Section 4.

This project has been funded by the European Union as part of the Horizon Europe Framework Program (HORIZON), under the grant agreement 101070109.

The **Dashboard LITE,**[4] shown in Figure 3, pursues a *linear, responsive* design approach to provide an intuitive tool for content analysis. It enables users to examine recent events and related online discussions across various Web-based and social media channels. As participants of the design workshops (see Section 3) stated the need for intuitive solutions that do not require extensive ex-ante training, the dashboard's LITE edition has been significantly revised to support a simpler content exploration workflow. This included a topic/metadata selector as a dropdown for accessing different analytic categories such as *sources, recency*, *sentiment*, *associations* or *topics*. Each selection represents a set of filters or search terms that a user can select to refine the results they receive.

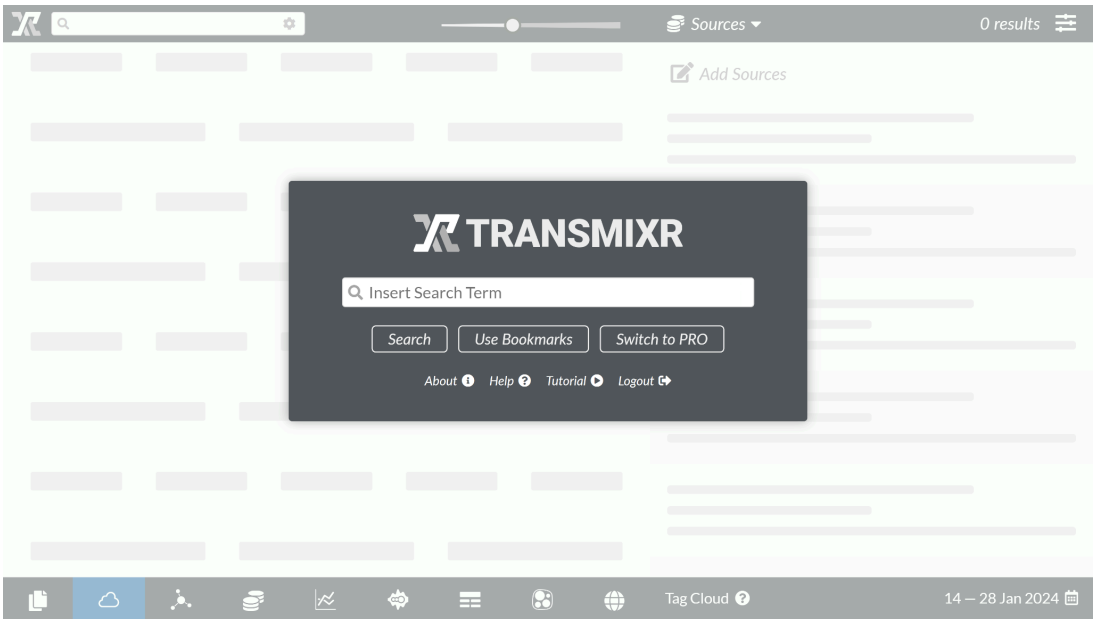

Figure 3: Landing Page of the LITE Version of the TRANSMIXR Visual Analytics Dashboard

The **Dashboard PRO** is more versatile and customisable, catering to professional users seeking in-depth analyses. As shown in Figure 4, it can be used to identify opinion leaders and predict forthcoming topics. It also incorporates impact metrics to assess the influence of specific sources and integrates emotion detection algorithms to offer insights into stakeholder perceptions regarding issues or events. TRANSMIXR introduced several UI enhancements to the *PRO Dashboard*. Users can now minimise individual sidebar sections instead of hiding them, offering a more customisable viewing experience. The export formats of the PRO dashboard have been improved, too, ensuring consistency in legends across different formats ("16:9 Slides" now include a "Story List", improved chart peak labels, revised icons, and the handling of little or no data in the reports).

[4] To assist consortium partners in testing and using the new features, a new video tutorial to introduce the features of the LITE version will be produced in Q2-2024.

This project has been funded by the European Union as part of the Horizon Europe Framework Program (HORIZON), under the grant agreement 101070109.

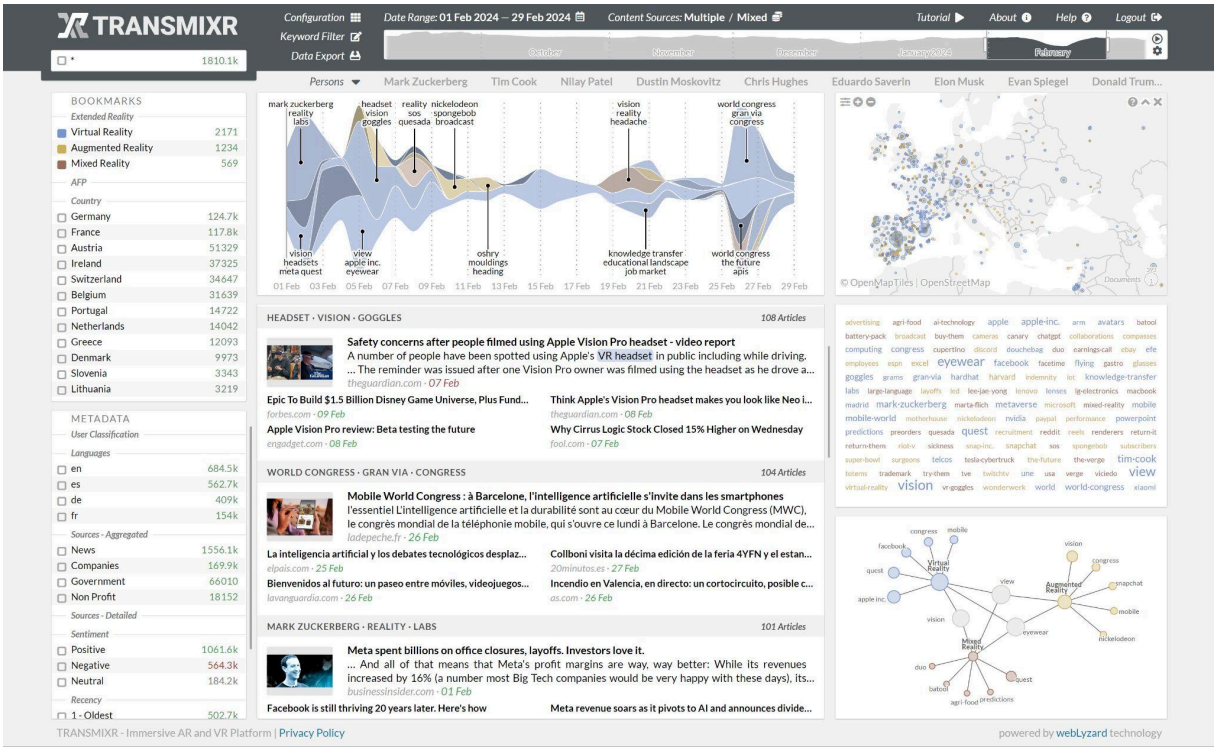

Figure 4: Landing Page of the Professional Version of the TRANSMIXR Visual Analytics Dashboard

Embeddable Widgets. The various dashboard visualisations are not only integrated into the automated *PDF Reports*, but also available as stand-alone components. As the *Storypact Editor* shown in Figure 5 (to be evaluated with AFP journalists from April 2024 onwards) demonstrates, visualisations can be integrated into content management systems or other Web applications hosted by project partners.

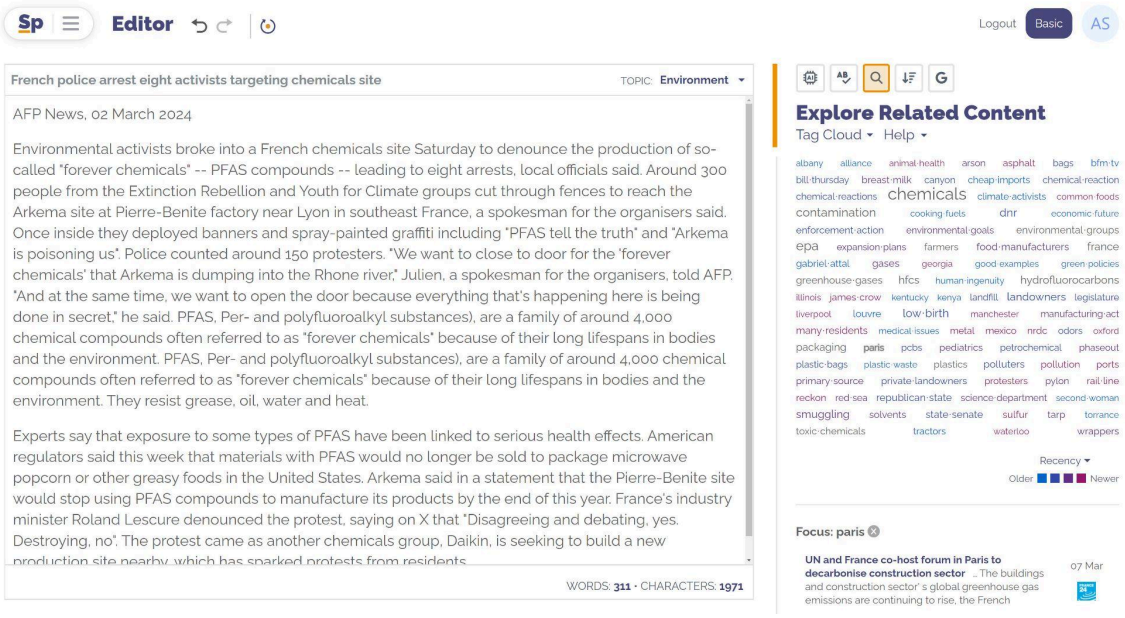

Figure 5: Screenshot of the Storypact Editor showing an AFP News article from 02 March 2024, including an embedded tag cloud from WLT to summarise associations and interactively explore related content

This project has been funded by the European Union as part of the Horizon Europe Framework Program (HORIZON), under the grant agreement 101070109.

3. Pilot Development

The TRANSMIXR project uses different use case pilots to evaluate its innovations. It explores the "Newsroom of the Future" in UC1 (news media with a focus on content production led by AFP) and UC2 (news media with a focus on content consumption led by TG4), while UC3 and UC4 target "performing arts" and "museum experiences", respectively. Following the four requirements elicitation workshops of WP1 with 15 end-users and 20 professional users, the initial dashboard prototype has been presented in three design workshops involving the use case partners. While the media-related use cases have received particular attention, the dashboard has also been discussed in the cultural heritage pilot, where WLT's content API represents a source for the pilot.

3.1 User Onboarding

Demos were organised in September 2023 between the webLyzard R&D team and journalists, and the project's media partners, presenting the different features of the dashboard and gathering feedback from the users. Given the initial focus on UC1 / content production, WLT distributed individual logins to the journalists who volunteered to test the platform. In collaboration with AFP, RTV SLO, TG4 and SPARK, feedback on how to best customise and extend the dashboard to serve as a news alerting tool for journalists that monitors emerging news in real time has been collected and translated into specific feature requests. Predictive analytics capabilities, for example, have been identified as a key function useful in newsrooms, helping to compile agendas of future events as candidates to be covered.

Demo Sessions. Three online demo sessions of the WLT platform were organised in September 2023 with the news partners of *UC1 - Understanding and Generative AI*. The objective of these sessions, which lasted about one hour each and were hosted by Arno Scharl from WLT, was to explain the major dashboard features and to analyse how they could be useful in newsroom workflows. The first two meetings took place on 15 Sep 2023: one with five AFP journalists, another one with two TG4 journalists, and the third one on 19 Sep 2023 with six journalists from RTV Slo. Among the main conclusions were that the dashboard could create value for newsrooms in two ways: (i) as a search tool to analyse trends and produce on-the-fly visualisations after further adaptation and alignment with the journalists' needs and (ii) as a real-time news monitoring tool to spot emerging stories and breaking news.

Design Workshops. Following the demo sessions, three design workshops were organised online with the different partners of the "Understanding and Generative AI" pilot (WLT, MOD, HSLU, CERTH, TCD, AFP, RTV SLO, SPARK). The first workshop on 22

This project has been funded by the European Union as part of the Horizon Europe Framework Program (HORIZON), under the grant agreement 101070109.

Sep 2023 focused on the news monitoring and research tool. The following two concentrated on the summarisation and formatting tools (CERTH) and the end-user experience. The results of the first design workshop are the most significant regarding the dashboard. The news partners put forward specific questions to WLT and MOD that emerged from the earlier demo session, including (i) how the platform could be extended for real-time monitoring and alerting, (ii) which sources the platform could ingest; e.g. which type of community-driven websites or major social networks, and (iii) how it could be integrated into journalistic workflows, e.g. as part of the newsroom content management system (CMS).

3.2 Requirements Elicitation

Following the design workshops, a second round of demo sessions of the TRANSMIXR dashboard took place with journalists from different offices of the AFP network (Berlin and Geneva as well as *Factstory*, the institutional branch of AFP). An online questionnaire was sent to the participants after the demonstrations to gather further requirements. The following examples illustrate the feedback gathered:

1 - Would the TRANSMIXR Dashboard be a valuable tool in your workflow?

- *Yes. If one can detect emerging news on social media in a timely manner.*
- *Yes. With its real-time monitoring, it could give me a hint of protests or major discussions going on in a country, or accidents/attacks that just happened.*
- *Yes. I would find it useful for analysing* the performance of an AFP story.

2 - Which features did you find most interesting?

- *The search and monitoring tools, and the cross-lingual aspect. The alert creation feature could be very useful and the predictive tool was interesting and clearly of interest for keeping the calendar up to date. I was also wondering if this tool could be used by AFP's graphics service to quickly generate graphs.*
- *Analysis of German media regarding the incorporation of AFP news.*
- *There are a lot of useful features, but the most interesting for my current tasks is to be able to rank the results with the "sentiment" filter.*
- *The ability to quantitatively and analytically record the success of our text products on various websites.*

3 - Which features are missing?

- *It would be helpful if texts could be translated directly within the programme, to avoid copy-paste into another AI translator. A broader range of social media monitoring would also be useful, including Telegram channels.*

This project has been funded by the European Union as part of the Horizon Europe Framework Program (HORIZON), under the grant agreement 101070109.

- *Tool to alert AFP of important developments in real time.*
- *Research of content on Twitter / show the keywords in the results / Connection with a generated AI to summarise what you can find in the results, not only one article but a global summary of what we find.*
- *I find the many existing options a bit confusing. Can I install a filter for routine use that shows me constantly the news I chose and need?*

4 - What would be an ideal monitoring tool for you?

- *Something like this would be good, but maybe a bit simpler - perhaps with less emphasis on analytics and diagnostics.*

3.3 Customisation Steps

It was decided to test the customised version of the TRANSMIXR dashboard in the Rennes AFP office, situated in Brittany, which covers the Western part of France. The dashboard is currently being customised with sources monitored by the journalists from the Rennes office, both local news sources and Twitter/X accounts. A "data access for researchers" application,[5] according to the Digital Services Act (DSA),[6] to access Twitter/X APIs is currently underway, which will be important for the newsroom pilot as Twitter/X remains an important source for many journalists. Once the Rennes dashboard is operational, the objective is to replicate the experience with other AFP offices in Berlin, Geneva, Warsaw, Madrid, and Paris. Ultimately, the objective is for each newsroom to be autonomous and able to configure the tool in a self-service mode in accordance with the region/topics it covers.

4. Dashboard Extensions

4.1 Multi-Color Storygraph

The *detection* (Nixon et al., 2019) and *visualisation* (Scharl et al., 2019) of emerging stories in the public online debate (= cluster of related documents, e.g. multiple articles reporting on the same issue or event) has been an important part of work done by MOD and WLT since the InVID (In Video Veritas) FP7 project.[7]

[5] algorithmic-transparency.ec.europa.eu/news/faqs-dsa-data-access-researchers-2023-12-13_en
[6] digital-strategy.ec.europa.eu/en/policies/digital-services-act-package
[7] www.weblyzard.com/invid

This project has been funded by the European Union as part of the Horizon Europe Framework Program (HORIZON), under the grant agreement 101070109.

Story detection not only identifies and characterises groups of related documents in digital content streams but also extracts metadata, evaluates its impact through temporal analysis of related articles, identifies keywords for summarising the content of a cluster, and visualises results via an interactive streamgraph, a type of stacked area graph (Byron and Wattenberg, 2008) to represent the temporal evolution of a dataset. The TransMIXR extensions have significantly improved the Story Graph's functionality and user experience with an enhanced visual representation, higher linguistic precision, and additional filtering capabilities described below.

The multi-colour extension addresses the need for real-time news detection as an entry point for journalists to explore emerging stories and their development over time. Previous versions of the *Story Graph*, as shown in Figure 6, could only render streamgraphs in a single colour, distinguishing stories through different opacity values but lacking the visual means to express additional metadata.

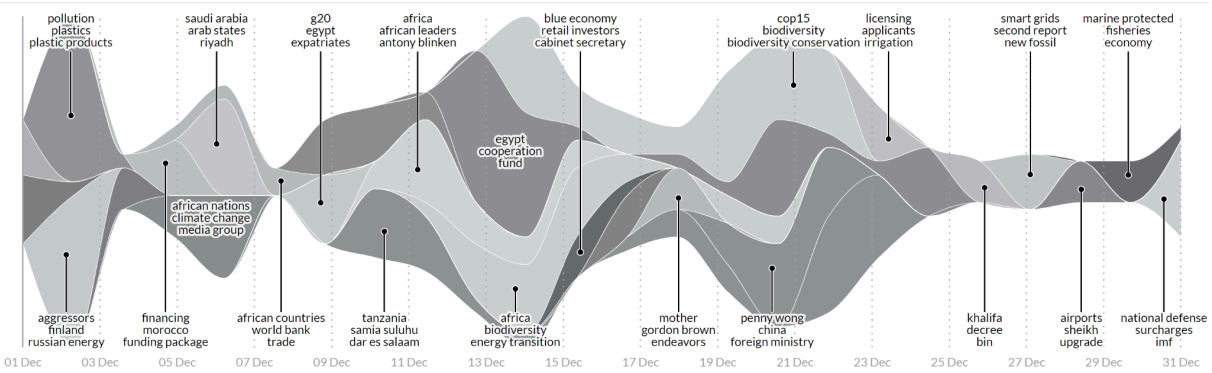

Figure 6: Unicolor streamgraph showing the evolution of stories in December 2022 on the topic "Sustainable Development Goals (SDGs), from WLT's SDG intelligence platform developed for the United Nations Environment Programme (UNEP)[8]

The new version (Figure 7) assigns colours based on various types of metadata. Special emphasis was on the flexibility of the implementation and the ability to customise the colour-coding process (and the underlying segmentation of the search queries). Stories that derive from multiple sources, for example, can be shown in a distinct colour. This facilitates tracking developments through visual cues and enables users to compare different topics, content sources, languages or other types of user-specified text classifications stored in the dashboard's "bookmark" system.

[8] www.weblyzard.com/unep-live

This project has been funded by the European Union as part of the Horizon Europe Framework Program (HORIZON), under the grant agreement 101070109.

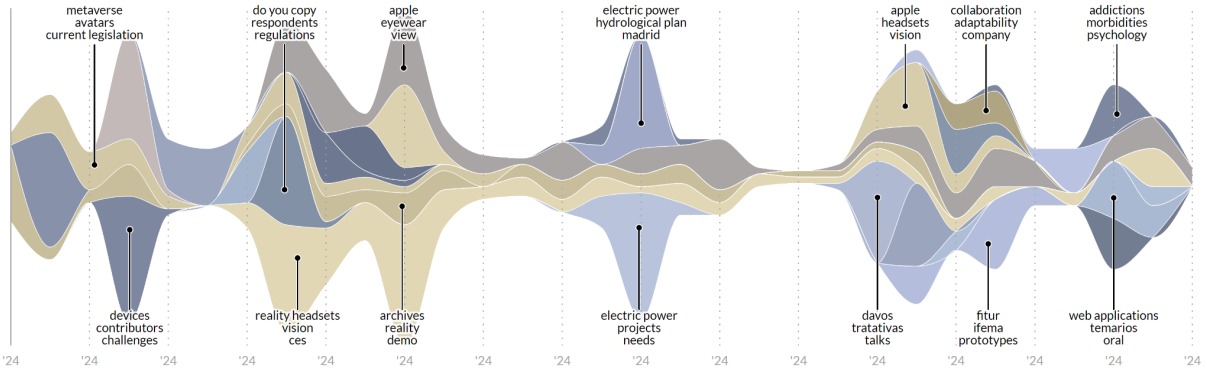

Figure 7: Multi-color streamgraph that reflects the platform's cross-lingual capabilities and shows the results of a query for virtual/augmented reality in English (yellow; n=1,213), Spanish (blue; n=1,186) and French (brown, n=204) articles from news outlets and the Websites of companies, non-profit organisations and government agencies.

In addition to the more flexible and faceted clustering process and colour scheme, several other aspects of the component have also been re-designed, focusing on both the quality of results and the responsiveness of the rendering process:

- **Clustering Algorithm.** Since effectively handling semantic overlap between clusters (e.g., documents appearing in multiple stories) remained a challenge of the previous Louvain community detection approach (Nixon et al., 2019), we have revised the time-sliced clustering algorithm with additional validation and filtering steps to reduce redundancy. This ensures that users are presented with a more consistent overview of specific narratives, which improves efficiency and the overall user experience.

- **Content Filtering:** Despite the algorithmic improvements described in the previous paragraph, keywords might be (correctly) annotated multiple times. To enable users to focus on specific narratives embedded in emerging stories and remove distracting elements, the Story Graph features a *keyword filter* where users can specify keywords not to be considered as labels for any of the identified stories.

- **Lemmatisation.** Complementing this, integrating a lemmatisation mechanism, a language processing technique to merge different term variations to a common base form, refines the story output by minimising repetition.

- **Response Time.** To enhance performance and responsiveness, the data loading mechanism has been optimised to handle multiple concurrent requests efficiently despite the additional complexity of underlying algorithms.

Given its importance for both the "Newsroom of the Future" pilot (UC1 / production and UC2 / consumption) and the tracking of evolving technologies, work on the story graph will continue in 2024.

This project has been funded by the European Union as part of the Horizon Europe Framework Program (HORIZON), under the grant agreement 101070109.

We will first experiment with different content metrics. Currently, the area of a cluster reflects the number of mentions. Additional insights might be possible by replacing frequency with other metrics such as *Share of Voice* (normalised for cyclic variations in overall news volume, e.g. fewer publications on weekends or during holidays), *Impact* (also considering the reach of a source; see above), or *Polarisation* using the standard deviation of sentiment to highlight contested issues.

A second set of experiments is planned to compare the currently adopted approach to computing cluster segmentation, based on the Louvain method for community detection (Nixon et al., 2019), with a dense vector embedding approach - assessing both the quality of the results and the achievable response time.

Status: Activated in December 2023; further development in progress.

4.2 Entity Inspection Tooltip

The existing tooltip of the WLT dashboard has been expanded as a new way to gain on-the-fly insight into an entity of interest (person, organisation, location or event) by simply hovering over it. In addition to detecting and presenting top documents associated with the entity and providing additional content on why certain terms are appearing in visualisations, the TRANSMIXR extension focuses on the display of additional metadata directly from the evolving knowledge graph, also referred to as Semantic Knowledge Base - SKB (Nixon et al., 2019).

The new entity tooltip feature, as outlined in Figure 8, provides additional context information and summarises background information on the object of interest. This feature is available when viewing an individual document with metadata analysis enabled, as well as in different visualisations that display entities (e.g., the *Entity List*). It is accessible via a second tab and presents short, pertinent information about the entity they clicked on, enhancing their understanding of the data displayed.

The tooltip includes an entity description, a thumbnail, a content filter with the operations that can be done with the entity (e.g., restrict or exclude certain topics), a set of thumbnails that point to related resources and links to the external resources with detailed information, and a list of all associated topics. Recent entity evaluation systems like Elevant (Bast et al., 2022) only showcase enough details to assess whether the entity predictions are correct. In contrast, our system provides additional context through the associations and list of resources, helping users easily identify the required entities. These tooltips help users examine the entities and reduce false positives (e.g., by signalling that certain entities are irrelevant in the context of a specific query), as it is well-known that they are quite common in search or named entity disambiguation tasks (Brasoveanu et al., 2018).

This project has been funded by the European Union as part of the Horizon Europe Framework Program (HORIZON), under the grant agreement 101070109.

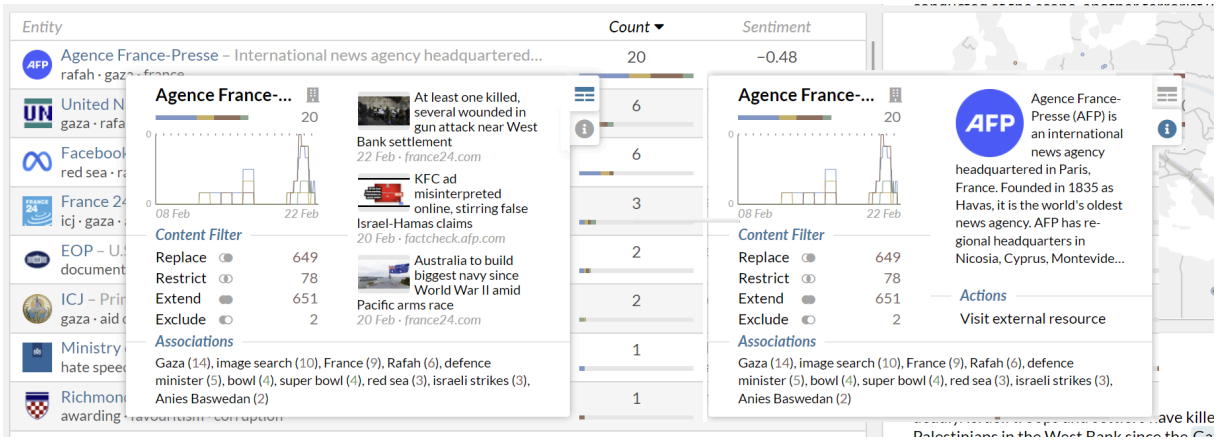

Figure 8: The two views of the Entity Tooltip. Left: Related documents. Right: Short information and link to the Semantic Knowledge Base (SKB) for additional background information on the entity.

Additionally, if the user has sufficient access rights, she can view the representation of the target entity in a separate application, the SKB Browser. Integrating more information than comparable tooltips surveyed by Hohman et al. (2020), the entity inspection tooltips are currently evaluated and will be included in the next major dashboard update.

Status: Final testing in progress, planned for a March 2024 release.

4.3 Document Highlighting

A new feature to highlight documents that will allow users to mark documents of interest to retrieve or export them later is complete and currently being evaluated. Adding and removing documents can be done by clicking the provided icon in the document list, which will mark it accordingly. A new checkbox will then be displayed in the sidebar, which acts as a way to quickly access the collected documents concerning the globally set filter, date restrictions and selected sources, similar to all other entries in the left sidebar sections. The checkbox can be selected to display the list of documents and create visualisations to gain further insights into the selected documents - including, for example, top keywords, locations or persons mentioned in the documents, etc.

Document highlighting is persistent across sessions so that work on the document collection can be continued later. A significant deviation from other elements selectable in the left sidebar is that the collection of highlighted documents can be directly exported via the Export Menu. The collection can be easily reset by opening the context menu, clicking on the topic in the sidebar, and clearing it.

This project has been funded by the European Union as part of the Horizon Europe Framework Program (HORIZON), under the grant agreement 101070109.

The selected documents could be integrated with the summarisation module reported in D2.1, creating a shortened representation of all selected documents at variable lengths. Coupled with a Generative AI service, the selected documents could also trigger the creation of additional text, images, or 3D/VR content related to the content of the highlighted documents.

Status: Completed; final testing in progress, planned for a March 2024 release.

4.4 Prediction Mode for Agenda Setting

The TRANSMIXR Dashboard offers a prediction mode that lists future events and anticipates which topics will be discussed on a given day, week or month (by using the referenced dates extracted from the document itself instead of its publication date for longitudinal analyses). The dashboard's predictive features are particularly useful for (1) agenda setting, optimising the editorial calendar, and identifying repurposing opportunities in the news pilot, and (2) identifying promotional opportunities for archived digital assets in the cultural heritage pilot.

The current limitations of the predictive features are twofold: the quality of the predictions and the level at which knowledge can be gained from the delivered documents. The quality depends on a critical mass of content. The XR-specific content feed (see Section 5.1), as well as the specific content feeds for AFP, are bound to provide a much richer layer of content and increase the usefulness and quality of the predictive capabilities.

One user reported that it is difficult to identify why a document was assigned a specific date in the future, which we will consider in upcoming interface design adaptations. To improve and speed up insight generation in prediction mode, we will emphasise referenced dates within documents that have been used for prediction in list-based views (documents, stories). This will ensure quick access to information on why documents have been matched and provide additional context.

Status: In progress; planned for June 2024

4.5 Real-Time Mode for Breaking News

To establish a near-real-time view, we will provide the means to activate a "live mode", where a new search request is sent off in customisable intervals, updating the results with the latest documents that were ingested and processed. The "live mode" will help users identify the latest developments. It will reuse and extend the existing module to colour-code results according to their recency - separating the search into four equally-spaced intervals (quartiles) based on the documents' publication dates, ranging from old to new documents, as shown in Figure 9).

This project has been funded by the European Union as part of the Horizon Europe Framework Program (HORIZON), under the grant agreement 101070109.

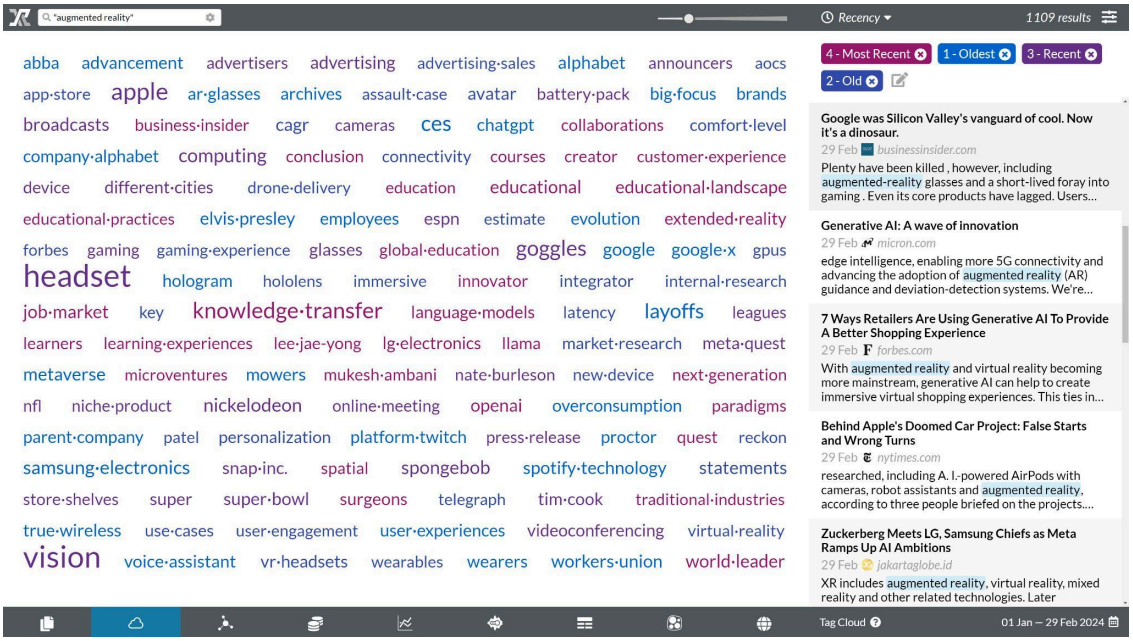

Figure 9: Associations with "augmented reality" in January and February 2024, colour-coded by Recency *(warmer colours indicate more current coverage), which will serve as the basis for the new "live mode"*

The "live mode" will use a more granular ensemble of colours to distinguish older and more recent topics as they emerge visually. To better highlight breaking news, the colours will not be allocated linearly. The most recent slices will be given a brighter signal colour for better visual separation from older timespans. We will develop user controls to determine how granular the periods should be calculated, i.e. separating the currently selected date interval into four (default) or a higher number of equally spaced periods.

Status: In progress; planned for June 2024

5. Content Metrics

The content metrics in the TRANSMIXR project include general affective indicators such as positive and negative *sentiment* and *polarisation* scores (WP2), as well as specific *impact indicators* such as *reach* and association with *desired* and *undesired* topics (WP6). Section 5.1 outlines the data collection process for a domain-specific XR content repository. Section 5.2 then describes the computation and consolidation of the various metrics.

5.1 Custom XR Content Feed

Typically, the content acquisition of WLT focuses on curated samples of Websites that are continuously expanded but are rather generic (e.g., news, public sector,

This project has been funded by the European Union as part of the Horizon Europe Framework Program (HORIZON), under the grant agreement 101070109.

NGOs, etc.), as well as partner-specific sets of relevant domain URLs (see Deliverable D2.1). Early in the project, it was determined that this hybrid approach would not suffice given the very dynamic domain of TRANSMIXR, where new AR/VR startups and emerging technologies (e.g., investors, consulting companies, etc.) lead to a very dynamic business ecosystem with a lot of previously unknown stakeholders, even to someone with insight into the market. To tackle this challenge, *Third-Party Web Indices* and *Knowledge Graphs* (Hogan et al., 2021) have been adopted as new content extraction approaches, introduced in the following subsections and complementing the content ingestion mechanisms outlined in D2.1. Both approaches used the following set of search terms to describe the domain of interest: *Extended Reality, Augmented Reality, Mixed Reality, Virtual Reality, Spatial Computing, Immersive, AR, VR*, and *XR*.

5.1.1 Third-Party Web Indices

This task performed time-sliced search queries using a third-party service (initial experiments suggested that the *Bing Web Search API*[9] would be most appropriate for this task) and follow-up metadata enrichment on a site level through WLT's Semantic Knowledge Base (SKB).

We collected Bing Search API data for July-December 2023 and aggregated it with WLT's content ingestion and enrichment modules, automatically backdating content based on their publishing date. A redundancy check ensured that content from sources already included in the continuously mirrored datasets was redirected accordingly. All remaining content was collected in a new repository. This repository was then used to identify the top publishing websites - about 2,000 domains in total - via the *Source View* component of the TRANSMIXR Dashboard (Figure 2, left). The results of this content ingestion process - 9,072 documents in seven languages (EN = 6,901, FR = 826, DE = 565, ES = 374, NL = 214, IT = 192) - were used for the follow-up processing steps. As the geotagged results in the geographic map (Figure 10, right) show, the referenced locations in the documents cover a significant part of Europe.

The Top 250 results were then exported into a spreadsheet for manual review and analysis. Additional metadata was extracted from the Wikidata KG[10] and Google KG[11] for each domain, including name, a longer description, country, location and logo and automatically added to the spreadsheet for a more extensive source profile. The added metadata simplified the following manual reversion that assigned a relevance ("very relevant", "relevant", "not relevant", and "not sure") to each source, where general technology news websites and gaming websites were considered as

[9] www.microsoft.com/en-us/bing/apis/bing-web-search-api
[10] www.wikidata.org
[11] developers.google.com/knowledge-graph

This project has been funded by the European Union as part of the Horizon Europe Framework Program (HORIZON), under the grant agreement 101070109.

"relevant" and as "very relevant" the ones that fit the XR/AR/ER focus. Additionally, large publishing or aggregation and indexing websites, e.g. *archive.org* or *springer.com*, that publish across many different topics, and large tech companies that operate in different sectors, e.g. *asus.com*, were flagged.

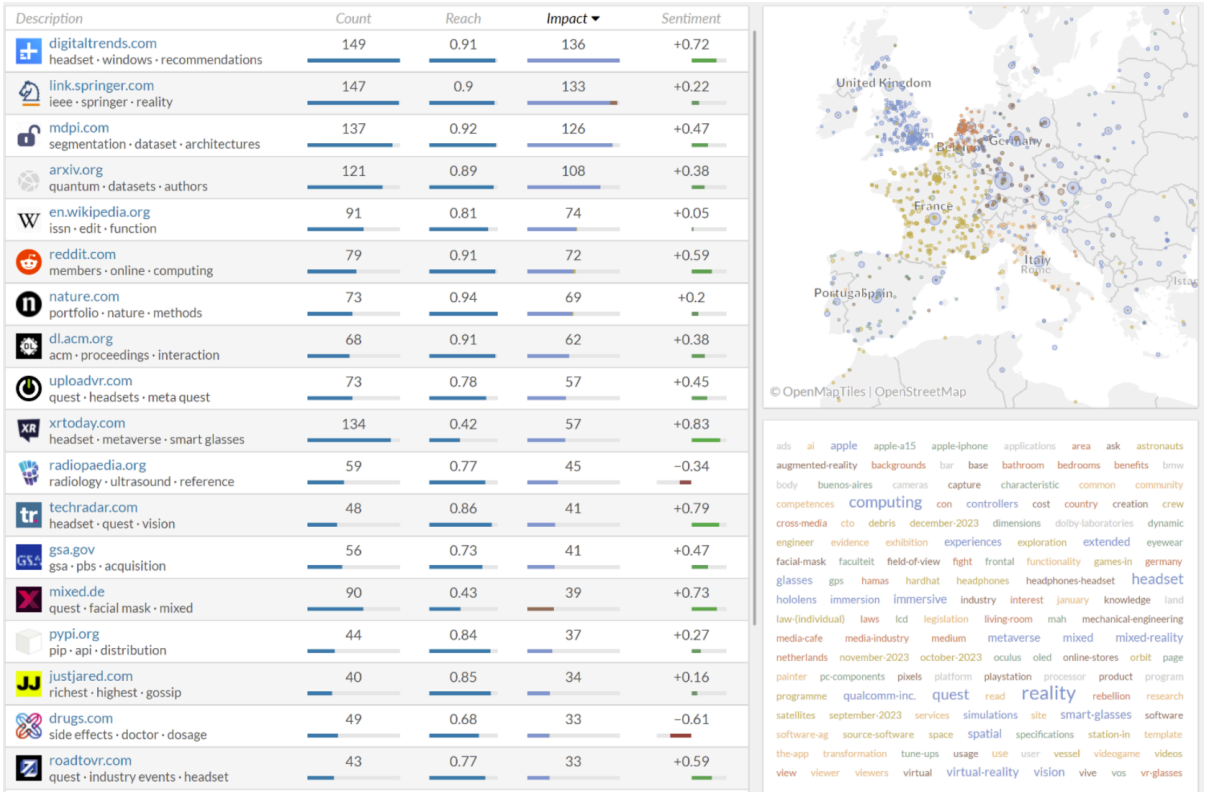

Figure 10. Aggregated view of immersive technology sources ingested via the BING Search API from English (blue), French (yellow), German (brown), Spanish (green), Dutch (orange) and Italian (light orange) sources, including a regional projection of mentioned locations and the major keyword associations by language

5.1.2 Third-Party Knowledge Graphs

Google Knowledge Graph data helped identify website sources relevant to the VR, AR and XR sectors as candidates for regular mirroring throughout the project. The Google KG API was addressed directly using the same search terms to retrieve KG entities (corporations, local businesses and persons) with a respective website domain linked. The metadata enrichment approach outlined in the previous section was performed, utilising Wikipedia to Wikidata links to receive full source profiles. With this approach, about 190 additional websites of various types of organisations were retrieved, as well as 180 persons. The person data represented an interesting experimental extension to identify persons of interest for a topic domain. They might be helpful for later phases of our research but are currently not the focus of the reported work. With this second approach, we identified a strong AR/VR focus within the game development industry.

This project has been funded by the European Union as part of the Horizon Europe Framework Program (HORIZON), under the grant agreement 101070109.

5.2 Metrics Computation and Consolidation

Digital media spaces are characterised by a culture of participation, where stakeholders discuss emerging stories, their perception of brands or the potential and the limits of new technologies. To compute appropriate metrics for assessing the impact of events or opinion leaders in the business ecosystem that TRANSMIXR targets, as well as the news coverage relevant to the use case pilots, we computed sentiment values, added MOZ reach metrics,[12] and normalised these metrics in a [0..1] interval. Computed keyword associations are a crucial building block of many of the dashboard's interactive visualisations.

Computation and Negation Processing. The WYSDOM[13] metric is a computational method to reveal the impact of global or regional events, shown as a stacked bar chart to show longitudinal patterns in conjunction with a radar chart to represent different aspects of an event. WYSDOM is a hybrid metric that contrasts semantic associations with (i) desired topics and the number of (ii) positive references with the number of (iii) negative references and the association with (iv) undesired topics. WYSDOM and the radar chart are part of the background knowledge available at the start of the project. The lack of real-time negation processing capabilities, however, limited the interpretability of their results. While sentiment and human emotions are pre-computed (Weichselbraun et al., 2022), the dynamic character of user-defined dimensions does not allow any preprocessing. To consider negated sentiment properly, we have been experimenting with different schemes to annotate sentences or specific phrases within sentences. For WYSDOM, the negated expressions could either be excluded from the metric calculation or shown below the x-axis. Another set of experiments is currently underway to test the impact of different segmentation strategies and how metrics computed on a sentence level, paragraph level, or document level could be integrated to improve the results obtained.

To progress the customisation of the indicators to be computed, a spreadsheet was shared with consortium partners to gather initial feedback on categories to assess ("immersive technologies", "innovation", etc.). For the immersive technology category, the initial list of desired terms contains terms such as *empathy, low-latency, real-time, photorealistic, captivating* and *motivating*. Undesired associations include *expensive, dizziness, headache, eye strain* and *nausea*. The category definitions will be expanded and refined based on automated suggestions and expert assessments. Once complete, the *desired vs. undesired* distinctions will be built into the dashboard as persistent bookmarks available to all users.

[12] moz.com/products/stat
[13] webLyzard Stakeholder Dialogue and Opinion Model; www.weblyzard.com/wysdom

This project has been funded by the European Union as part of the Horizon Europe Framework Program (HORIZON), under the grant agreement 101070109.

Content Filtering and Monitoring. We are working on validating and fine-tuning the results to reduce noisy data contained in the ingested content - for example, offers from web shops just selling AR/VR headsets, and to improve the filtering and metadata enrichment steps. Valuable new sources (e.g., startups or small- and medium-sized companies developing new immersive technologies), will be added to our continuous monitoring, creating a dedicated market watch sample to compute the T6.3 impact metrics and provide feedback for the TRANSMIXR exploitation efforts in T6.4.

6. Outlook and Conclusions

Deliverable D2.2 has presented the current status of the content metrics and dashboard development activities of the TRANSMIXR project. To ensure the reliability and stability of new components added to the application, including the *Story Graph* and other visual elements, TRANSMIXR follows a multi-step approach consisting of (i) comprehensive unit tests, (ii) thorough internal testing by the development team, and (iii) iterative feedback loops with test users from use case partners (in this reporting cycle, the refinement focused on the news industry and the requirements of partner AFP). WLT and SPA plan to provide alternative delivery formats to expand the range of possible applications and unlock additional exploitation opportunities, for example, in the form of browser extensions.

The user feedback from the pilots (see Section 3) regarding workflow support will be incorporated by developing a topic-sharing functionality, as well as a feature to bookmark or rate specific articles that a journalist might want to revisit while working on a story. We plan to have the initial prototypes of both features available for consortium partners to test in time for the M18 review. Additional features planned for 2024 include a sidebar section for *IPTC Newscode*[14] classifications (Sayed et al., 2023), customisable *keyword management*, allowing users to choose the types of associations to display as labels across the dashboard (e.g., persons, organisations, locations, keywords), and the asynchronous delivery of data exports and automated reports, enhance the user experience of the PRO dashboard and supporting new stand-alone applications (e.g., a regular newsletter).

While the initial feedback iterations focused on the “Newsroom of the Future” pilots, additional extensions to support other pilots will be tackled from Q2-2024 onwards. This will be accompanied by the activation of the daily Web crawling of the sources identified via the XR content feed, as well as the ingestion of descriptions of cultural heritage objects - e.g., videos, sound recordings or 3D/XR assets from sources such

[14] www.iptc.org/standards/newscodes

This project has been funded by the European Union as part of the Horizon Europe Framework Program (HORIZON), under the grant agreement 101070109.

as *NISV*[15] and *Europeana*.[16] The ability to analyse a specific source or to contrast the coverage from different channels or stakeholders will help users understand how media narratives are constructed and how the same story can be presented in various ways.

7. References

Bast, H., Hertel, M., & Prange, N. (2022). ELEVANT: A Fully Automatic Fine-Grained Entity Linking Evaluation and Analysis Tool. In *Proceedings of the 2022 Conference on Empirical Methods in NLP: System Demonstrations* (pp. 72-79).

Braşoveanu, A., Rizzo, G., Kuntschik, P., Weichselbraun, A., & Nixon, L. J. (2018). Framing named entity linking error types. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)* (pp. 266-271).

Byron, L. and Wattenberg, M. (2008). Stacked Graphs – Geometry & Aesthetics. *IEEE Transactions on Visualization and Computer Graphics*. IEEE Computer Society. 14 (6): 1245–1252.

Hohman, F., Conlen, M., Heer, J., & Chau, D. H. P. (2020). Communicating with Interactive Articles. *Distill*, *5*(9), e28.

Hogan, A., et al. (2021). Knowledge Graphs. *ACM Computing Surveys*, *54*(4), 1-37.

Nixon, L., Fischl, D., Scharl, A. (2019). Real-Time Story Detection and Video Retrieval from Social Media Streams, In *Video Verification in the Fake News Era*. Eds. V. Mezaris et al. Basel: Springer. 17-52.

Sayed, M. A., Braşoveanu, A. M., Nixon, L. J., & Scharl, A. (2023). Unsupervised Topic Modeling with BERTopic for Coarse and Fine-Grained News Classification. In *International Work-Conference on Artificial Neural Networks* (pp. 162-174). Cham: Springer Nature Switzerland.

Scharl, A., Hubmann-Haidvogel, A., Goebel, M., Schaefer, T., Fischl, D. and Nixon, L. (2019). Multimodal Analytics Dashboard for Story Detection & Visualization, In *Video Verification in the Fake News Era*. Eds. V. Mezaris et al. Basel: Springer. 281-299.

Weichselbraun, A., Steixner, J., Braşoveanu, A.M.P., Scharl, A., Göbel, M. and Nixon, L.J.B. (2022). Automatic Expansion of Domain-Specific Affective Models for Web Intelligence Applications, *Cognitive Computation*. 14(1): 228-245.

[15] www.beeldengeluid.nl/en/knowledge/assignment-or-talk/archive
[16] www.europeana.eu/en

This project has been funded by the European Union as part of the Horizon Europe Framework Program (HORIZON), under the grant agreement 101070109.