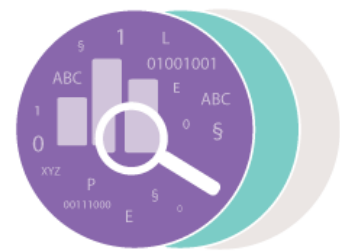# FutureTDM
Explore . Analyse . Improve

REDUCING BARRIERS AND INCREASING UPTAKE OF TEXT AND DATA MINING FOR RESEARCH ENVIRONMENTS USING A COLLABORATIVE KNOWLEDGE AND OPEN INFORMATION APPROACH

Deliverable D2.5

Workshop summary report 2

Project

Acronym:	**FutureTDM**
Title:	Reducing Barriers and Increasing Uptake of Text and Data Mining for Research Environments using a Collaborative Knowledge and Open Information Approach
Coordinator:	SYNYO GmbH

Reference:	665940
Type:	Collaborative project
Programme:	HORIZON 2020
Theme:	GARRI-3-2014 - Scientific Information in the Digital Age: Text and Data Mining (TDM)

Start:	01. September, 2015
Duration:	24 months

Website:	http://www.futuretdm.eu/
E-Mail:	office@futuretdm.eu

Consortium:

SYNYO GmbH, Research & Development Department, Austria, (SYNYO)

Stichting LIBER, The Netherlands, (LIBER)

Open Knowledge International, UK, (OK/CM)

Radboud University, Centre for Language Studies The Netherlands, (RU)

The British Library Board, UK, (BL)

Universiteit van Amsterdam, Inst. for Information Law, The Netherlands,(UVA)

Athena Research and Innovation Centre in Information, Communication and Knowledge Technologies, Inst. for Language and Speech Processing**, Greece, ** (ARC)

Ubiquity Press Limited, UK, (UP)

Fundacja Projekt: Polska, Poland, (FPP)

Deliverable

Number:	**D2.5**
Title:	**Workshop summary report 2**
Lead beneficiary:	**OK/CM**
Work package:	WP2: INVOLVE: stakeholder integration, workshops and advocacy
Dissemination level:	Public
Nature:	Report (RE)
Due date:	31.05.2017
Submission date:	29.05.2017

Authors: **Freyja van den Boom**, OK/CM

Contributors:
Burcu Akinci, SYNYO
Gwen Franck, LIBER
Stefan Kasberger, OK/CM
Lieke Ploeger, OK/CM
Martine Oudenhoven, LIBER

Review: **Olga Jurkowska**, FPP

Table of Contents

Table of Figures

Table of Tables

1 INTRODUCTION

The second FutureTDM workshop was held at the European Parliament in Brussels on the 29th of March 2017.

The purpose of the thematic, multi stakeholder workshops is to provide the opportunity for stakeholders to discuss drivers and barriers to the take up of TDM and to feed into the core deliverables of the FutureTDM project.

This report provides a summary of the proceedings, the main insights taken from the discussions, the feedback and the media coverage of the workshop. Looking back the workshop has achieved its goal in providing a platform for the project to present the results and receive feedback from the stakeholders involved.

1.1 Organization, location and timing

The main organization of the workshop was in hands of LIBER together with the different consortium partners. The workshop was kindly hosted by leading copyright MEP Catherine Stihler at the European Parliament. Having the workshop again in the heart of the EU ensured we would reach a wide number of stakeholders involved in policy and TDM. The date was chosen to fall within the timeframe for informing the copyright reform debate within the lead EP Committee.

We were unfortunate that Brexit got triggered on the day of the workshop and that our initial host Catherine Stihler was not able to join us due to a last minute schedule change.

Figure 1: Screenshot Twitter

1.2 Agenda

The FutureTDM Workshop Agenda was set up as follows.

> *Registration*

14.00-15.00 Session One - TDM Findings and Recommendations

© 2017 FutureTDM | Horizon 2020 | GARRI-3-2014 | 665940

Welcome from the Chair
Dr Paul Ayris, *Co-Chair of the League of European Research Universities (LERU)*

FutureTDM the policy context, the project and next steps,
Melanie Imming, *Projects Manager, LIBER (Association of European Research Libraries)*

The Economic Opportunity
Jan Strycharz, *Economic Analyst, Foundation Project: Poland, Digital Center*

FutureTDM's Recommendations
Marco Caspers, *Research Associate, Institute for Information Law, University of Amsterdam*

Questions from the floor and summation from the Chair

15.00-15.15 short break

15.15-17.00 Session Two - Improving TDM in the EU with Targeted Policy

Introduction to sector specific recommendations - key issues
Kiera McNeice, *British Library[1]*

Workshop - feedback from TDM stakeholders

- *Allied 4 startups*
- *Digital Europe*
- *OpenMinted*
- *Jozef Stefan Institute*

1.3 Attendance

Invites were sent to the FutureTDM network including participants from previous events and those who have expressed interest via the website, social media etc. We used an online sign-up system that would ensure personal details were protected. All MEPs were sent invites via the internal EP system thanks to the office of Catherine Stihler. LIBER also sent invites to MEPs working specifically on copyright and parliamentary advisors. All partners were asked to promote the workshop within their own extended networks.

We closed external registration 2 weeks early because of the room capacity for session 2. We set up a waiting list for those who wanted to register afterwards. We were able to accommodate everyone who asked to be on the waiting list. This system worked well - we met the room capacity at the event. In total, we had 53 participants in session one and 38 in session 2.

Diversity and Stakeholder representation

We were very pleased to have members from the relevant stakeholder communities present including representatives from policy, advocacy organisations and representatives from the Parliament Commission and Council, industry, academia and members from the general public. The following

---

[1] http://www.bl.uk

Figure 2 shows the gender diversity amongst the workshop participants and their affiliation covering all relevant stakeholder communities.
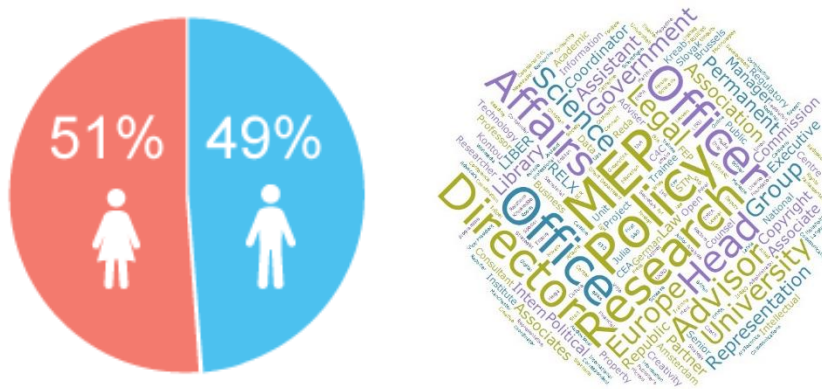

Figure 2: Gender diversity % of attendees and word cloud with affiliations

2 THE WORKSHOP SESSIONS

The workshop was split into 2 sessions – the first one was a formal panel looking at the project findings and focused on its two most recent reports on economics of TDM and policy recommendations.

In Session 2, we held a stakeholder round-table, where the issues from session one were discussed and the project outlined how we will be developing guidelines for increasing TDM uptake.

2.1 The Panel

The event was moderated by Dr Paul Ayris, Co-Chair of the League of European Research Universities (LERU).[2] He began by stating that the issue of TDM is of great relevance to LERU's members and that he was pleased to be working together with the FutureTDM consortium for this event.

Dr Ayris explained the purpose of the workshop which is to hear more from the Commission funded FutureTDM project on ways to improve uptake of TDM in the EU.

Figure 3: FutureTDM workshop Impressions: Paul Ayris explaining the purpose of the workshop

---

FutureTDM, the project and next steps

Figure 4: FutureTDM workshop Impressions: FutureTDM Workshop II panel at the European Parliament

Melanie Imming, EU Projects Manager, LIBER (the Association of European Research Libraries) gave an overview of the FutureTDM activities. Imming started out with an introduction to 'text and data mining', after which she gave an overview of the core project activities and the most recent reports generated by the project. One of the main achievements of FutureTDM is that it generated some much-needed connectivity on the 'human' level - connecting and activating multiple stakeholders to increase awareness, skills and legal knowledge around the issues related to TDM. She stressed that everybody with lawful access should be allowed to do TDM.

The Economic Opportunity

Jan Strycharz, Foundation Project: Poland, Digital Center, followed with the project's research on the economic value and impact of TDM. He started by pointing out that a lot of innovation is focused on 'big data' and related issues (such as processing speed and storage), but TDM is about much more than that: it is also about analysis, about human capacity building and training. Strycharz stressed the potential return on investment in TDM.[3]

---

[3] http://www.futuretdm.eu/knowledge-library/

Figure 5: Twitter screenshot slides FutureTDM presentation

FutureTDM's Overarching Policy Recommendations

Marco Caspers, Research Associate at the Institute for Information Law, University of Amsterdam talked about potential legal barriers to TDM. The main impediments are fragmented database law across Europe generating uncertainty for practitioners and companies, who will not engage in any potential illegal activity. Removing this uncertainty, by introducing a mandatory exception for TDM is crucial to create legal certainty for TDM practitioners. [4]

---

[4] https://www.slideshare.net/FutureTDM/futuretdm-workshop-ii-29-march-74335338?ref=https://www.slideshare.net/FutureTDM/slideshelf

Figure 6: Twitter screenshot presentation slide on legal barriers and principles[5]

Feedback and questions from the audience

After the panel presentations, the audience raised questions - mainly about the position of Europe in relation to other markets such as the US. Lack of a clear legal framework and of trained professionals who have the skills to conduct TDM might become an impediment for European actors to be able to participate and compete in a global market.

Figure 7: FutureTDM workshop Impressions: session one

2.2 Workshop session 2

Paul Ayris, (LERU) chaired the second panel. This panel consisted of a combination of experts from the FutureTDM consortium, the OpenMinTeD project and our board of advisors as well as representatives from the TDM stakeholder communities to contribute perceptions, barriers and opportunities from the different perspectives.

FutureTDM 'developing guidelines'

Kiera McNeice, FutureTDM Project Officer at the British Library, explained that in the next project phase the insights gained throughout the project will feedback into developing recommendations and proactive solutions to support greater uptake of TDM. The next steps include putting together a series of guidelines focusing on specific group of stakeholders to help them overcome the barriers we've identified throughout the project.

These guidelines along with the projects policy recommendations and best practices will be disseminated and promoted in a variety of ways. The main deliverables will be hosted on the FutureTDM website, and key actions will be summarized and promoted as separate awareness sheets, through blog posts and our knowledge base.[6]

Figure 8: The FutureTDM awareness sheet collection

The project has already highlighted some areas where we feel that TDM uptake could be improved through stakeholder guidelines and they are illustrated in our tree of digital knowledge. The idea of these guidelines will be to give concrete, actionable advice to fill some of the gaps we've identified in support for TDM. The guidelines will be developed as part of an ongoing process, with as much iterative feedback as we can get our hands on.

---

[6] http://www.futuretdm.eu/awareness-sheets/ and http://www.futuretdm.eu/category/blog/

The guidelines so far:

Legal Issues: We have addressed and will continue to address legal uncertainty by providing clear, accessible information about what people need to be aware of when carrying out TDM

> So, for example if you're a startup without access to expert legal advice, you can quickly educate yourselves in the fundamentals, and get an idea of whether you're exposing yourself to any risk.

Licensing: We provide clear explanations of what licenses actually allow (including open CC licenses), and what people can reasonably expect to be included in a bespoke license

> So that if you are for example a university offered a license by a publisher, you can understand whether that license will allow your researchers to carry out the TDM work they would like to do.

Data management: The aim is to educate people on the different technical requirements that apply to bulk access to content for the purpose of TDM, as opposed to individual access to content, and how people can make their data better available for TDM

> For example, if you create or store content in a repository, what metadata should you be including so that your content is genuinely re-usable for TDM?

Policy guidelines for universities: We aim to encourage universities to support TDM through both their research and education arms.

> For example, by helping university senior management understand the needs of researchers around TDM, and potential benefits of supporting it

Building the TDM community: We will broaden the TDM user-base, by demonstrating applications of TDM in areas that aren't traditionally thought of as data-driven.

> We will develop and share brief case studies on a variety of different TDM applications, promoted with awareness sheets and via our knowledge base online.

Reflecting on these issues, our chair Mr Ayris brought up the issue of awareness and preparedness among researchers in making their data ready for TDM. A recent UCL survey found that many researchers don't think about research data management early enough (or at all) in the process of their work.[7] Often data is still stored on paper and a significant number are not aware of what TDM is. So, we are at the absolute beginning of academic appreciation of the power of TDM and the potential scientific and economic benefits – this will need to be addressed.

We then heard from three members of the FutureTDM Expert Advisory Board:

---

[7] http://discovery.ucl.ac.uk/1540140/1/UCLRDMSurvey_report_Feb2017.pdf

The FutureTDM Advisory Board

Damir Filipovic, Director Digital Enterprise and Consumer Policy at DIGITALEUROPE highlighted that for global tech companies and SMEs legal and licensing issues will be of most interest.[8] Additionally, content access is also important. For these stakeholders, the scope of planned exception should be expanded. In terms of research, collaboration is important and the private sector (who often provide technical tools) should not be left behind when it comes to reform for innovation.

Figure 9: Impression of speakers

Lenard Koschwitz, Director European Affairs at Allied for Startups talked about how Allied for Startups have been conducting a range of workshops, meeting entrepreneurs from across the continent and that there's an overwhelming interest in TDM.[9] Startups are not looking for a free ride, but are willing to acquire licenses. In the majority of cases however, the data is free and legal, so TDM should be allowed. Current wording of exception for scientific research is valuable to create a safe space, but at the same time the EC risks creating an unsafe space for those who don't fall within the scope (such as startups), and they remain in legal uncertainty. The current formulation is even limiting universities, for example. A big motivation for many technical universities is to encourage successful spin offs – this is under threat if the exception is not widened. Startups will look for the best ecosystem for them. If the EU is not dealing with TDM properly, startups might decide to go and look for another ecosystem (outside of the EU).

---

[8] http://www.digitaleurope.org/
[9] http://alliedforstartups.org/

Figure 10: FutureTDM workshop Impressions

Marko Grobelnik, Text Mining Researcher at the Jozef Stefan Institute emphasized that data has always been around, but now is much more accessible and available for technological improvement.[10] This throws up questions around legal issues of privacy and ownership and uncertainty about whether you are allowed to use data that has been accessed. In addition, there's a huge talent gap and the needs of the market are much bigger than current HR supplies. The EU is not bad at it, but we are losing the talent we have to the 'cool companies' in the US. We also don't seem to manage to import talent from elsewhere.

OpenMinTeD

To explain more about the work that our sister project OpenMinTeD carries out, **Sophie Aubin** from INRA gave an overview of the best practice the project are offering, including TDM tools and services. She explained how the platform will integrate existing resources facilitating digital standardization and interoperability. This should also help provide legal clarity.

Discussion

In the discussion that followed with the workshop participants we had a lively debate covering various issues including:

● *The importance of recognizing data protection and privacy*
● *Whether the scope of the exception should encompass both text and data*
● *Skills being given the same attention as technology*
● *The need for more practical examples on how TDM is working*
● *How to use 3rd party tools and platforms*
● *competition vs copyright and what should be regulated where*
● *Licenses under the exception*
● *TDM for different levels of experience from young people to management*

---

[10] https://www.ijs.si/ijsw/JSI

- *Encouraging investment where there is no legal certainty*
- *Whether we need to change our mindset in the EU – do we need an 'innovation exception'?*

Figure 11: FutureTDM workshop Impressions

3 DIGITAL TREE OF KNOWLEDGE DISCUSSION

At the end of the session we invited participants to contribute to our digital tree of knowledge, writing their organization or sector on the leaf stickers and adding them to the branch seen as most important to them. The feedback on the design and concept was very positive.

Perhaps unsurprisingly, the legal guidelines branch gathered the most leaves. The info from these discussions plus the feedback forms/tree will feed directly into Deliverable 5.3.
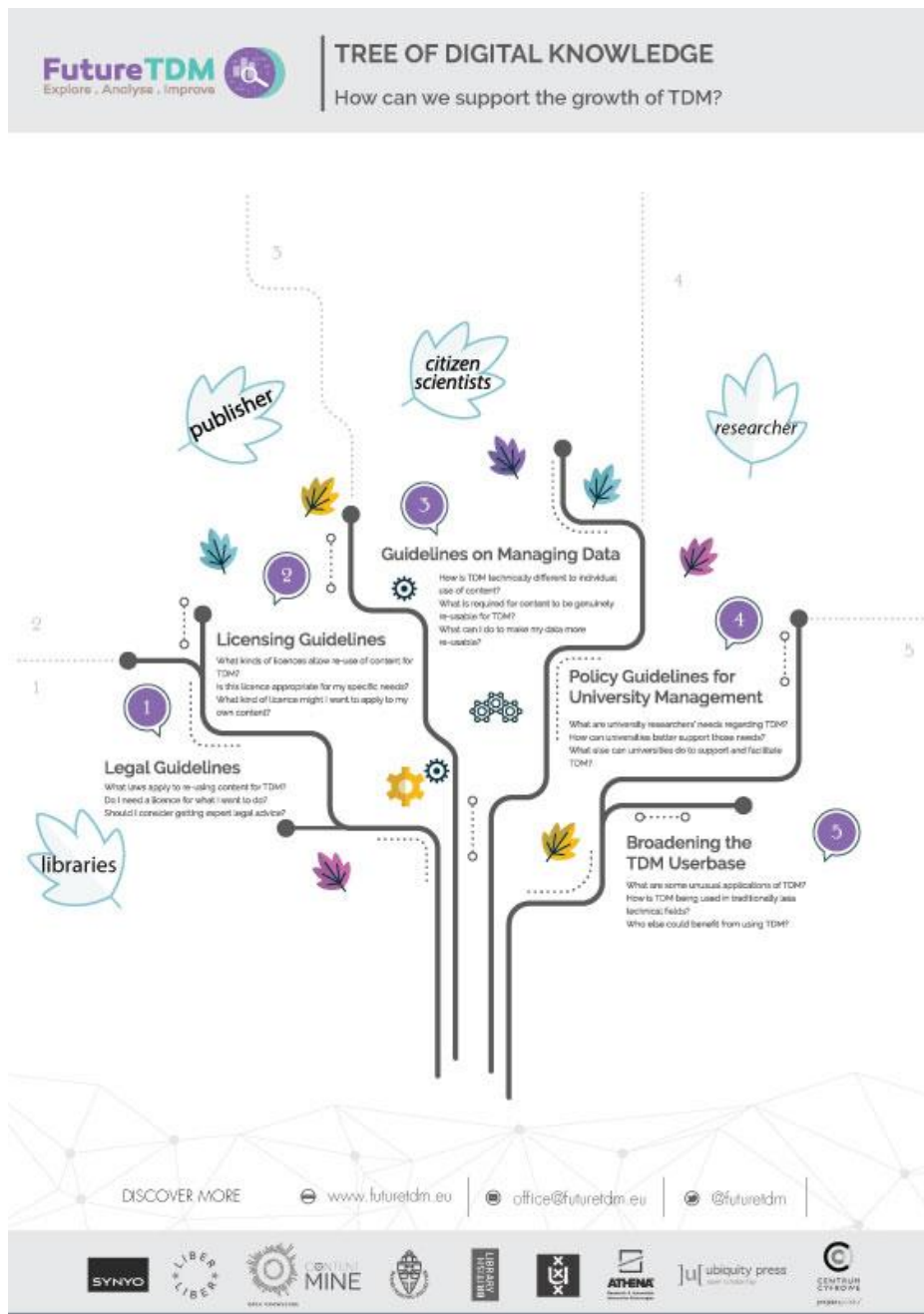

Figure 11: Discussion in session 2 focused on our tree of digital knowledge poster

4 FEEDBACK

As it is vital for FutureTDM to receive input from its stakeholder communities, different means were used to engage with the TDM community. The workshop participants were invited to share their sector-specific views at the workshop or online and to comment and/or suggest solutions to help improve the uptake.
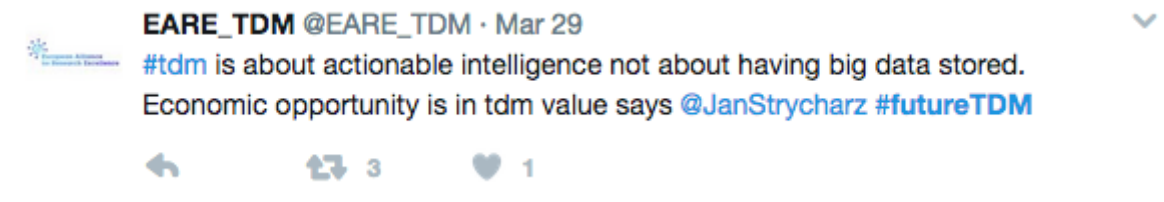

Figure 12: FutureTDM workshop Impressions

Feedback cards

We provided feedback cards for all participants to encourage everyone to provide input. We received a variety of very useful responses to be taken into consideration moving forward with our research.
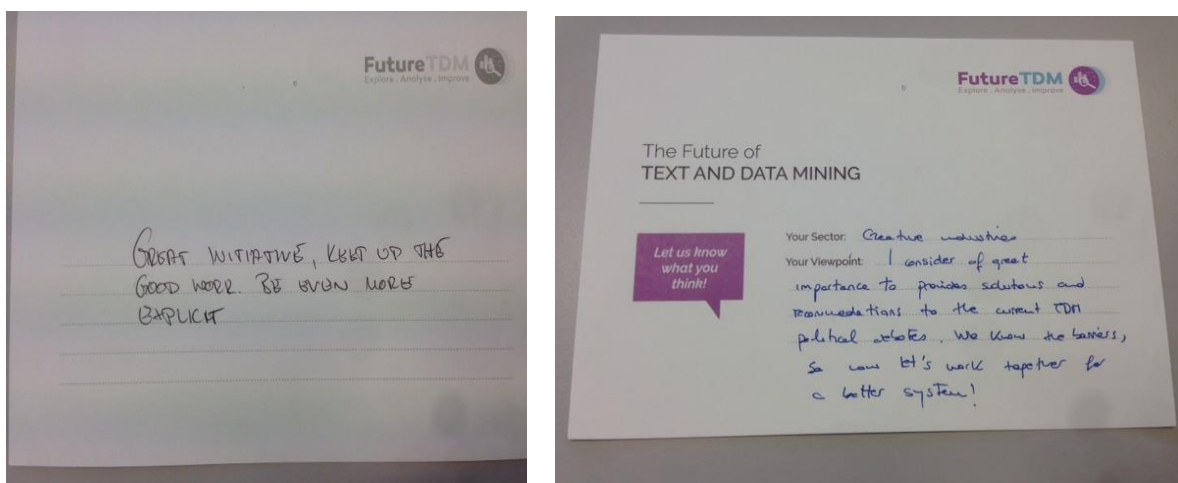

Figure 13: Feedback cards

TABLE FEEDBACK CARDS	
Sector	**Feedback [edited]**
Technology	Great initiative and project. FutureTDM is a great voice to have in the current copyright debate.
Law	A specific TDM exception is ok for now, but in my opinion in the long run we need more fundamental changes (e.g. non-consumptive use of copyright-protective works; extended consent for personal data processing, anonymization standards…)
Scientific research	There must be <u>no restriction on subject</u> (like scientific content in France) as any text is potentially a study object.
Natural sciences, research	No publicly funded research and results under private ownerships.
Legal research	The lawmaker should put more emphasis on solving technical issues.
Market research and data analytics	Would be great to have some actual examples of TDM. Try to bring in the 'commercial' (non-academic) actors as well.
Business	Potential legal barriers for entrepreneurs and start-ups need to be removed, or at least decreased, in order to enhance TDM.
Creative industries	I consider of great importance to provide solutions and recommendations to the current TDM political debates. We know the barriers, so now let's work together for a better system!
Think tank - digital	More practical focus is needed. The discussion is very abstract to be approachable (legal standardization, semantics etc). The legal report and the copyright act should also be wider explained, as this is among the first times when the argument enters the legal EU debate.

Table 1: Responses to feedback cards

5 DISSEMINATION AND MEDIA IMPACT

The workshop provided multiple opportunities for dissemination including the following

- **Blog posts**
  The FutureTDM consortium partners provided live coverage of the event using Twitter which resulted in an overall increase in Twitter presence. Shortly after the event we published a dedicated blog on the website covering the day.
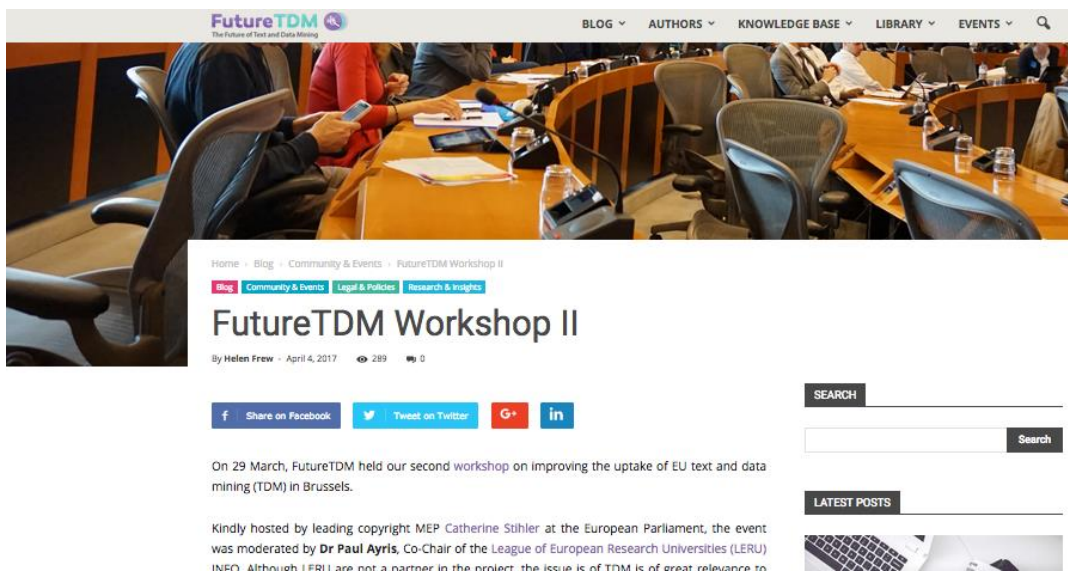

Figure 14: screenshot FutureTDM webpage featuring blog

- **Video Interviews**
  During and after the workshop the speakers were invited to present their view and answer a few questions to be recorded for our video channel. All video's are uploaded on our Youtube account.
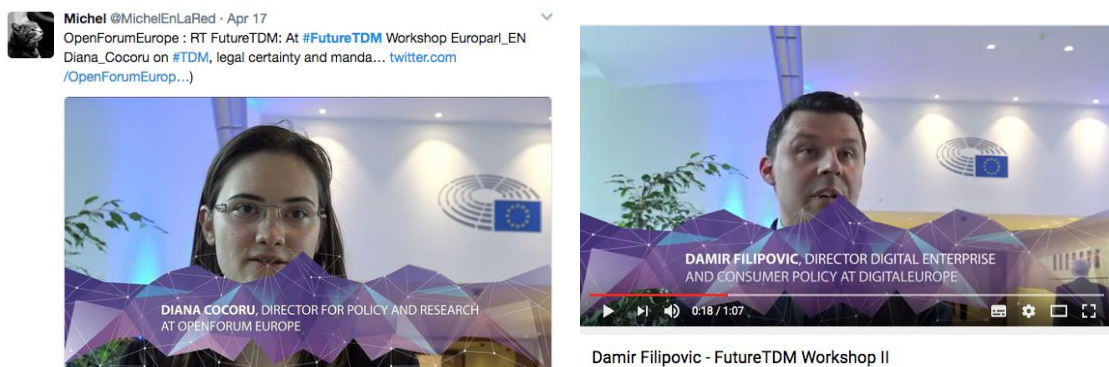

Figure 15: Screenshot social media

- **External coverage**
  The workshop received coverage from different stakeholders who attended the workshop.

ALLIED
FOR
STARTUPS

Wind up your clock, Helen

- **Brexit: Are countdowns exciting if they end in 728 days?**
- **Copyright: More time to take startups in the equation.**
- **Scaleups & MEPs: An opportunity to visit Silicon Valley.**
- **TDM or better TM + DM?**

#Copyright4Startups:
klab Berlin helps classrooms go digital. Founder Benjamin believes that protecting IP shouldn't come at the expense of innovation. Here's what he said about the Copyright proposal.

Your weekly Tech-Countdown from Brussels:

728 days to go: the government of the UK submitted its Article 50 letter to European Council President Tusk on Wednesday. One news network promptly responded with a **Brexit** clock. For two years, the UK will negotiate its exit from the EU. While it'd be a fool's errand to predict what the results might be, we're asking whether PM Theresa May has startups in mind?

Startups require an ecosystem that provides access to talent, capital and a regulatory environment that encourages innovation. London's Silicon Roundabout has thrived in the last years and is clearly Europe's biggest hub, especially in FinTech.

TDM Special:

Future **TDM**'s workshop in Parliament showcased how data analytics helps Europe's economy grow. Globally, Text and Data Mining will be create a value of €17bn by 2020, more than half of it €10.3bn can could come from the EU if rules allow for it.

Copyright: The deadline for amendments in the Committee on Legal Affairs was pushed back by a week. This means, see what startups have to say and read our position paper!

Engine

Member's Corner: Engine's report analyses whether online filtering could work in practice. The answer is straightforward: Even the best techniques won't be able to make complex judgements on copyrighted content. Are there lessons to be learned for EU lawmakers?
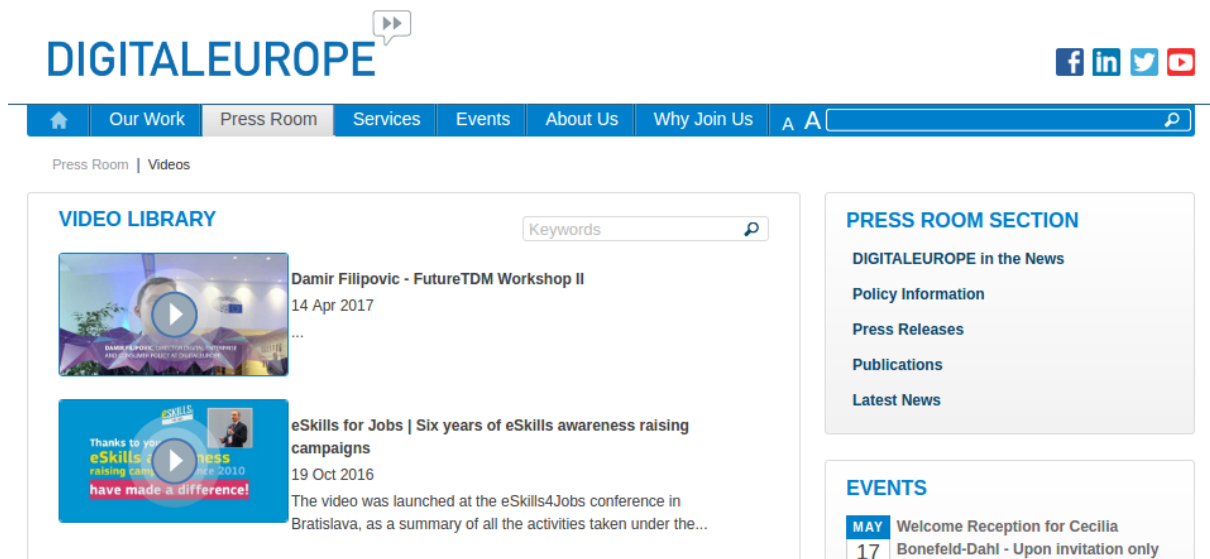
Figure 16:Allied for Startups newsletter, 31 march 2017

Figure 17: Video interview posted on DigitalEurope website

Analytics and insights

Twitter

We posted tweets during the whole day and workshop attendees and speakers posted or retweeted these and other relevant tweets on twitter. The following screenshots of the FutureTDM twitter account analytics of the day of the workshop and the days after show an impact on all levels.[11]
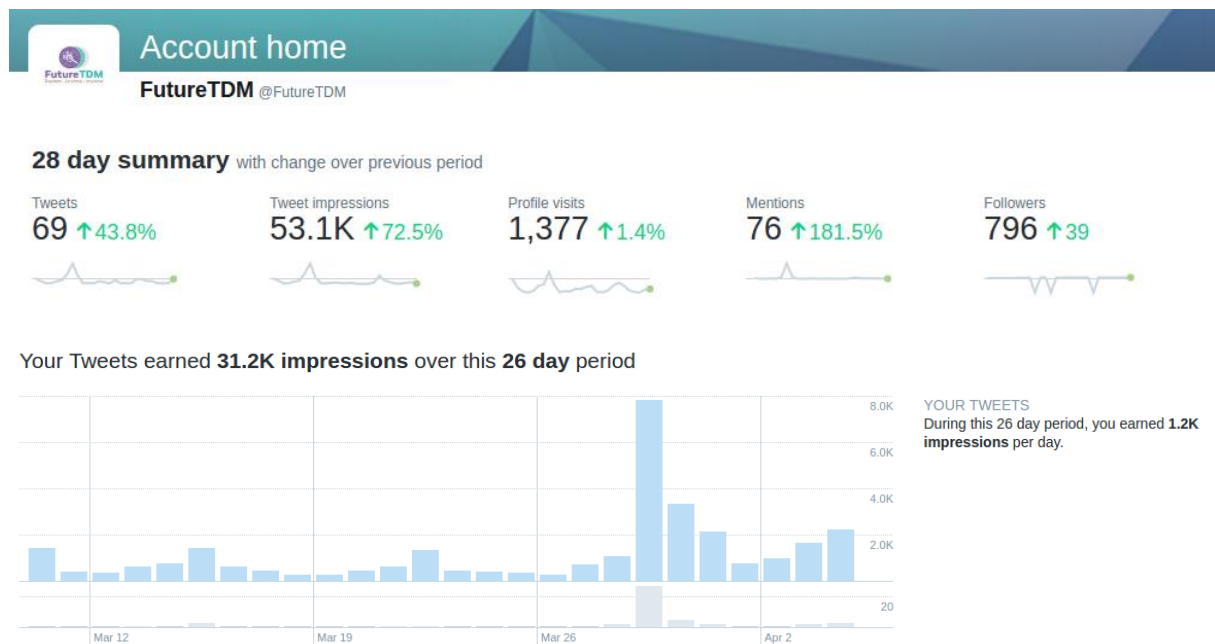

Figure 18: Twitter Analytics

---

[11] https://twitter.com/FutureTDM

Since the workshop, we have had an increase in profile views, mentions and people signing up to follow our account. See figures below.
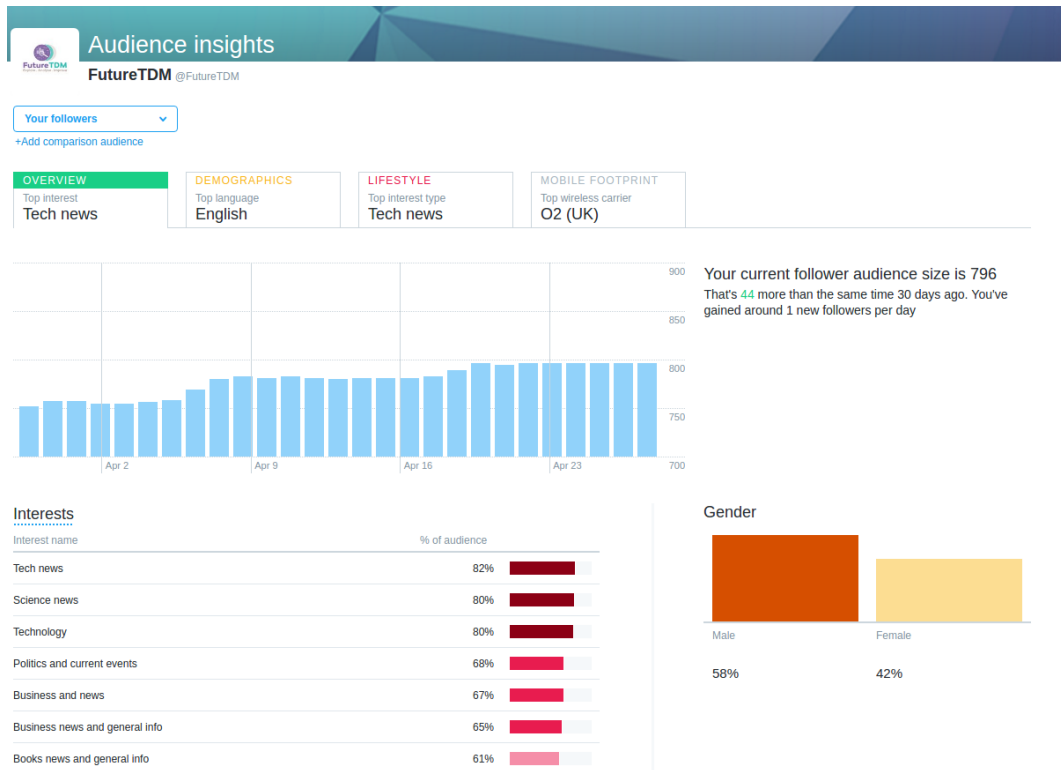

Figure 19: Twitter Audience insights

FutureTDM platform website

The FutureTDM website also received an increase in the amount of visitors as is shown in the following figure:
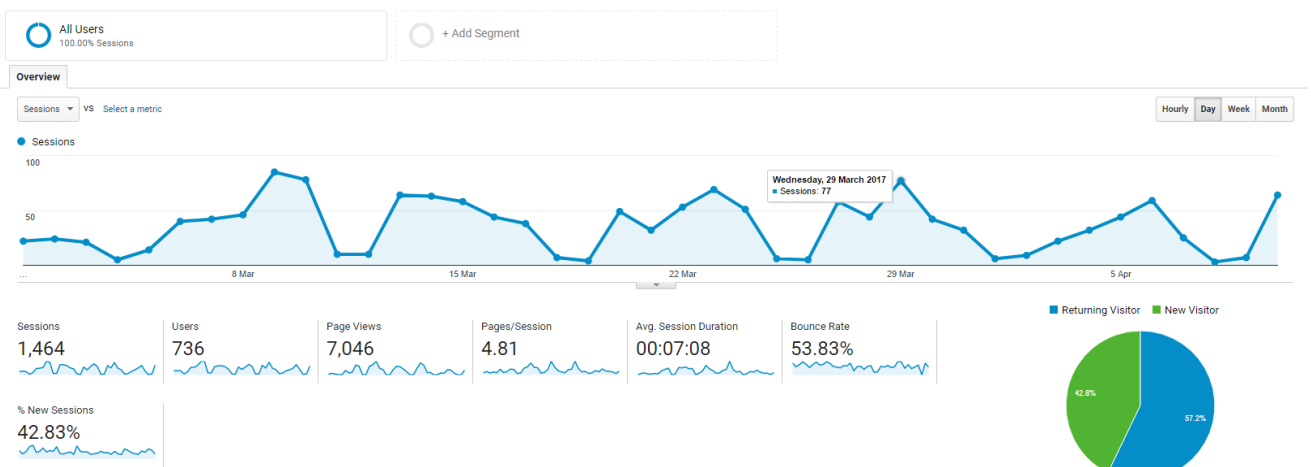

Figure 20: FutureTDM platform page view

6 CONCLUSIONS

This report provides a summary of the second iteration of the multi-stakeholder workshops on barriers and enablers for TDM. Where the first workshop's focus was on the barriers to the uptake of TDM as identified in WP2. The second workshop, which is the topic of this report, looked more at the enablers for TDM e.g. the FutureTDM policy principles and guidelines.

We can conclude that the workshop has been successful in its purpose of providing an opportunity for the TDM stakeholder community to discuss the results of the FutureTDM research as well as the uptake of TDM more generally. We reached our target amount of participants to include our main stakeholder focus groups. We invited and welcomed representatives from policy, academia, research and industry as well as people from the general public to join us.

All contributions, comments and feedback received will feed in to the final phase of the FutureTDM project core deliverables of WP3, 4 and 5 and the Roadmap for the uptake of TDM in WP5.