

# Guidelines for Supporting TDM at Universities

## 1. Introduction

Text and data mining can refer to a broad range of different activities, but fundamentally it involves using computer algorithms to analyse content and generate new knowledge. Computers are able to process data on a much larger scale than any human reader, and assess more variables across more datasets. The potential benefits to researchers are vast – but uptake and use of TDM technology in Europe is thus far lagging behind other areas of the world.

The FutureTDM project has produced several sets of guidelines to help address specific challenges to greater use of text and data mining (TDM) technologies. These include guidelines to help practitioners better understand the legal and licensing situation around TDM, and guidelines to help creators of data and content manage and share their data. In each case these guidelines are intended to be relevant to all stakeholders involved in TDM and data creation activities, from academia to industry to the general public. One finding that emerged from the FutureTDM project, however, was that universities in particular have a key role to play as stakeholders in the TDM landscape. These guidelines aim to highlight just how and why universities play such a crucial role, and provide practical advice on what universities can do strategically to support TDM within their own institutions and across the wider TDM landscape.

As we will explain, supporting TDM and related skills<sup>1</sup> has the potential to bring a wide range of benefits to universities and their students and researchers.

## 2. Universities as key stakeholders

It would be difficult to exaggerate the potential impact universities can have on the uptake of TDM technologies. In FutureTDM's policy framework and recommendations, we identified universities as influential stakeholders in all stages of the TDM value chain.<sup>2</sup>

### 2.1 The TDM value chain

#### Content creation

Universities and their researchers generate large amounts of research data and other content that are potentially valuable to TDM practitioners. As discussed in our Data Management Guidelines, making sure this content is managed and shared according to best practices offers benefits for universities both as content creators and owners, as well as potential re-users of others' content.

#### Content dissemination

Universities are increasingly playing a role in the dissemination of content created by their staff and researchers. Again, by following best practices for data management and sharing, universities can maximise the visibility and value of their shared content.

#### Text and data mining

---

<sup>1</sup> Foundational skills that are relevant to TDM include general data literacy, best practices in data management, and data science and analytical skills. These are collectively described as data- or TDM-related skills and practices throughout these guidelines.

<sup>2</sup> See the [FutureTDM Policy Framework](#) as outlined in [Deliverable D5.1](#) (PDF) for a detailed discussion of the TDM value chain, and an overview of the roles various stakeholders can play in supporting the uptake of TDM.

The use and development of text and data mining technologies is a rapidly growing field in universities around Europe. Whether in dedicated departments or research groups, or as part of broader programmes of data science and related technologies, many applications of TDM technology are happening within and around universities.

## Value creation

Similarly, universities play a key role in the translation of TDM research into new knowledge, insights and technologies. Whether this be through disseminating the results of TDM research projects, or turning newly-developed TDM tools and algorithms into spin-out companies, universities play a key role in generating scientific and economic value through the use of TDM.

## Skills and education

Although not an explicit part of the TDM value chain, all aspects of the TDM landscape rely on stakeholders and practitioners having access to the necessary skills, education and support to carry out TDM projects. Universities play a crucial role in supplying the necessary skills, education and awareness that underpin all other aspects of the TDM value chain.

## 2.2 The role of universities

As well as their direct involvement in the above aspects of the TDM landscape, the structure of universities, and their positioning in wider social and economic landscapes, make them uniquely placed to support the uptake of TDM technologies:

- As cross-disciplinary bodies, universities can facilitate the sharing of knowledge and skills across multiple domains, helping to bridge the gaps between areas that are typically more advanced in data-related skills, and those that are less likely to work with large-scale digital data.<sup>3</sup>
- As institutions, universities negotiate licences for access to content on behalf of their researchers, and can use their bargaining position to ensure those licences allow TDM.<sup>4</sup>
- Universities are trusted voices in the community, and have large networks through which they can communicate and demonstrate the value of TDM technologies and their applications.
- Through public-private partnerships with industry, universities can ensure that students and researchers learn data-related skills that are relevant and valuable to industry and the economy.

Improving awareness and support for TDM and related skills and technologies in universities will therefore bring further benefits beyond just the aspects of the TDM value chain in which universities are directly involved.

## 2.3 Benefits to universities

The principal reason for universities to invest resources in supporting TDM is that data science and analysis is fast becoming fundamental to all areas of research and education. Students need to understand data and how to manipulate it in order to be comfortable and confident in the modern world. In the words of one professor at a leading university, “Data science is the new IT.”

Education in data analytics is fast becoming essential, even in what may seem to be unlikely fields. In the fashion industry, for example, business is largely conducted online – which means students need a sensitivity and

“Data science is  
the new IT.”

<sup>3</sup> See section 14 of FutureTDM Deliverable D3.3+ for a detailed report on the skills gaps across different disciplines in TDM education and skills.

<sup>4</sup> For guidelines on how appropriate licences can help support TDM, see our Guidelines for Licensees.

awareness of the value that data analytics, including TDM, can bring to their business models.

More broadly though, universities benefit from the aggregated impact of their researchers,<sup>5</sup> and TDM has the potential to increase the progress of research exponentially. From helping researchers to sort through the ever-growing volume of academic literature, to developing entirely new research and analysis techniques based on new algorithms, TDM can on the one hand save time and money, and on the other become a foundation for all kinds of innovations and research discoveries. Supporting and encouraging training in TDM and related skills will make university graduates better prepared to operate in a big data world. TDM technologies will help to make research more effective and efficient, uncovering new insights and saving time and money through automated processes. We should not be content to leave the teaching of TDM and related skills to just a few specialised universities, when the potential value to all research fields and sectors of the economy is so significant.

### 3. Challenges

Through interviews, workshops, and other consultations with stakeholders, the FutureTDM project identified several significant barriers that hinder greater uptake of TDM in universities. These are discussed in depth in the FutureTDM report on policies and barriers to TDM in Europe,<sup>6</sup> but the major barriers are summarised below.

#### 3.1 Lack of awareness

Even though TDM technologies began emerging in the 1990s, TDM is still seen by many as a new, or even a “niche” field. Particularly outside of traditionally data-driven disciplines, there can be less awareness of or interest in TDM – despite its vast potential for applications in all areas. However, awareness does vary by institution, and in some cases digital humanities programmes are already developing and implementing applications for TDM.

#### 3.2 Fragmentation of resources

In the context of supporting TDM, the sheer scope and breadth of activities that universities typically engage with are potentially one of their greatest strengths, allowing them to bridge gaps between different fields and roles. But actually overcoming those gaps can equally be one of the greatest challenges to implementing new policies to support TDM.

Depending on a given university’s structure, there may be multiple groups interested in, or even already pursuing, TDM-related initiatives and ideas. This can lead to confusion and duplicated effort – but most significantly, without a central or coordinated approach to supporting TDM, progress and learnings are not shared among different parts of the university.

This fragmentation is a significant barrier to the uptake of TDM, and makes it difficult for anyone interested in TDM to know whom to turn to for support or advice – whether they be a researcher looking to develop TDM applications themselves, a researcher looking for others to help apply TDM solutions, or even a librarian looking to better understand TDM technology. Time and again, the feedback from stakeholders consulted during the FutureTDM project was that the absence of coordinated support from their universities forces them to rely on ad hoc personal networks for information and support regarding TDM.

#### 3.3 Skills gaps

A further challenge is bridging the gap between students or researchers in a given domain, and experts in data-related skills. Just as researchers may not have the expertise to understand TDM tools and applications,

---

<sup>5</sup> [The LERU Roadmap towards Open Access](#), June 2011 (pdf)

<sup>6</sup> FutureTDM Deliverable D3.3+

experts in TDM may not have the domain-specific knowledge to understand how those tools can be applied in a given discipline. As one university librarian commented, “You almost need two specialities,” to be able to understand how TDM can be utilised to solve problems within a given discipline.

### 3.4 Lack of resources

The lack of awareness, and fragmentation of existing interest in TDM, has obvious repercussions in a broader lack of resources supporting TDM and related skills. Without funding for dedicated roles to support TDM, universities are left to rely on staff with the ability and motivation to support TDM in their own time, on top of other responsibilities.

This is of course difficult for existing staff to manage on top of their existing workloads, as supporting TDM not only requires learning about these new and emerging technologies, but also coordinating among different stakeholders who may need to be consulted for a TDM project, and advocating for further uptake of and support for TDM. There is a crucial need for roles that have the time and backing to properly support and coordinate TDM activities across universities, as well as staff with a range of capabilities who can bridge the gaps between domain experts and TDM experts.

## 4. Paths forward

As of April 2017, few universities have taken concrete steps towards designing or implementing strategic policies to support TDM. However, a large number of the stakeholders we talked to expressed a desire to do more to support TDM within their institutions, as well as an interest in learning from any progress other universities have made.

In this section, we suggest some steps you or your institution might take to work towards developing and implementing policies and strategies to support data science and analytics. We heard from many stakeholders that there is still considerable foundational work to be done around awareness and implementation of data literacy, data management, and data science before addressing support for TDM specifically; the example cases below therefore focus largely on supporting these foundations of TDM. However, the general principles in each step should apply to supporting TDM and related skills at any stage.

More detailed case studies will be available on the FutureTDM website.<sup>7</sup>

### 4.1 Demonstrate need

In 2015, Ghent University carried out a survey and series of interviews across all research and education faculty to understand what skills the university library, and the institution as a whole, should focus on investing in and supporting. Among other things, the results of this project highlighted a need for better skills in data management and data science.

The university library used the results of this project as evidence to investigate how education in data management skills could be introduced into core university curricula. Teaching of some of these skills has now been implemented in doctoral schools, and ultimately the library’s project aims to make these part of the curricula of every Master’s degree by 2018 (with the depth and detail covered varying by field).

Although not explicitly educating students about TDM itself, laying the foundations for good data management practices is a key first step towards supporting greater use of data analytics. More generally, consulting with the university community and gathering evidence to support a need for better investment in data-related skills is a strategy that can be used to drive policies of investment and support in this area.

---

<sup>7</sup> <http://www.futuretdm.eu/awareness-sheets/>

## 4.2 Involve all stakeholders

As discussed above, within a university there are likely to be many people and communities who stand to gain from better investment in data-related skills, or who would be impacted by policies in this area. Without buy-in from all these stakeholders, it is difficult to gather support for new strategies and policies around data science and analytics. These stakeholders may include researchers in all fields, from sciences to humanities; librarians and library staff; heads of faculty in education, as well as research; students; IT departments; and other specialised support groups for activities such as software development and e-research.

Several universities have described planning or holding workshops to bring together stakeholders from libraries, IT departments, and research and education faculties to ensure that all parties' needs and concerns are understood. In some cases, these have been supplemented by ongoing working groups, to continue evaluating and refining plans for policies around data. These sorts of initiatives to “get everyone on the same page” are key to developing strategies to support data science and analytics.

## 4.3 Understand relevant university processes

Each university has its own organisational structure, and its own internal hierarchies; it would not be possible for these guidelines to suggest a general strategy for introducing data-related skills to education curricula or specific research departments. It is therefore a crucial first step to consult and communicate with education and research faculties to understand how their processes work, in order to devise a centralised, coordinated plan to integrate better support for TDM and data science into those areas.

Looking again at Ghent University, the university library initially had little insight into how the education arm of the university was structured or organised. Through consultations with the directors of research and education of every faculty of the university, they developed an understanding of the “learning lines” followed through degree structures, and the basic competencies these cover. While some competencies were discipline-specific, many were general skills taught across all degrees. By identifying places where skills around data literacy and management might fit into existing competencies, the library was able to build a case to suggest places to incorporate these skills into existing education streams – and also to highlight gaps in existing “learning lines” and competencies.

In general terms, understanding how education policy and practice is structured at your university will help you to understand and express how education in data-related skills might be implemented, in a way that education policy-makers understand and recognise.

## 4.4 Consolidate information about resources

As discussed above, support for different aspects of TDM, where it exists, is often fragmented across multiple departments of a university. Many people are unsure of whom they can turn to for help within their own institutions, and what resources might be available to them. Bringing together information about and access to these resources in a centralised, coordinated place makes it much easier for anyone interested in TDM to understand whom they can turn to for support.

This applies to more than just the technical aspects of carrying out TDM itself; researchers may need advice on how to manage, store and share their data, or expert legal advice to understand whether their intended use of TDM is lawful. Researchers should be able to quickly and easily discover how their libraries, IT departments, legal teams, and any other relevant departments can help support TDM.

## 4.5 Identify early adopters and promoters

As with any emerging technology, there are likely to be people within your university community who are already interested in TDM and related skills.

The University of Cambridge’s “Data Champions”<sup>8</sup> are a group of volunteers who are given support and freedom to host workshops, and create and disseminate materials around data management best practices. The Data Champions programme began with a call for volunteers from within the university’s research community to help support data management. These self-identified Data Champions bring with them the domain-specific knowledge and expertise necessary to understand what data management means in practical terms within their specific disciplines, and as researchers themselves, they also have a trusted voice among their peers and colleagues.

The Data Champions programme has thus far seen significant progress in encouraging awareness and best practices around data management. Identifying and working with researchers who are keen to understand and promote the importance of data-related skills appears to be an effective step towards bridging the gap between researchers and data experts.

#### 4.6 Introduce appropriate incentives

Where funding is available, this can be used to hire experts for dedicated roles to support TDM and related skills. For example, the Delft University of Technology, with support and funding from senior management, has recently begun a hiring process for “Data Stewards” to help support best practices in data management at the university. If you are considering a programme of support for TDM, it is worth considering funding sources both internal and external to the university.

However, other incentives can also be used to encourage people within the university community to support TDM and related practices. At the University of Cambridge, volunteer Data Champions are given public recognition via the university website,<sup>9</sup> as well as internal recognition in the form of reference letters to their heads of department. At University College London, PhD students in computer science have been working to create data science education materials aimed at a pre-university level, and have been happy to volunteer their time towards this project as a public good.

The best way to understand what incentives would encourage stakeholders within the university community to help support TDM and related practices is, of course, to consult with them.

#### 4.7 Share your results and progress

Thus far, few universities have taken concrete steps towards policies supporting TDM. Many more are interested in supporting TDM in principle, but uncertain how to go about developing and implementing policy in this area. If you or your institution has had success in developing strategies around supporting TDM and data-related skills, we encourage you to share and promote your success stories with others.<sup>10</sup>

### 5. Summary of key points

Universities are uniquely placed to support the uptake of TDM, but to fully capitalise on this potential ultimately requires a central, coordinated approach to bring together information, resources, and people in this field. Some ways that you can gather and encourage support and awareness around TDM at your university include:

- carrying out surveys or interviews to demonstrate a need for TDM support;
- involving and engaging with all stakeholders across the university;
- understanding the organisation of education and research faculties, and how TDM and related skills might fit into existing processes;

<sup>8</sup> <http://www.data.cam.ac.uk/intro-data-champions>

<sup>9</sup> <http://www.data.cam.ac.uk/datachampions>

<sup>10</sup> We welcome everyone to share their experiences with TDM on the [FutureTDM blog](#).

- bringing together information about and access to resources, to make these easily discoverable by people interested in TDM;
- finding and identifying early adopters who are personally interested in promoting or pursuing TDM;
- introducing incentives for people to engage with TDM-related initiatives; and
- sharing experiences and success stories so that others can learn from them.

Case studies of how particular universities have begun to address support for TDM will be shared via the FutureTDM blog<sup>11</sup> and awareness sheets.<sup>12</sup>

---

<sup>11</sup> <http://www.futuretdm.eu/category/blog/>

<sup>12</sup> <http://www.futuretdm.eu/awareness-sheets/>