**Outline for the Supplementary Information**

This supplementary file, describes more in details the work titled:

**TAPS Responsibility Matrix: A tool for responsible data science by design**

This work was performed from 2019-2022

For an up-to-date version of TAPS-RM please refer to the project repository for the TAPS-RM tool - github.com/MaastrichtU-IDS/responsability-matrix-survey

**Introduction:**

The development of the Transparency Accountability Privacy and Societal Responsibility Matrix (TAPS-RM) was prompted by the need for a framework guiding ethical practices in the rapidly evolving field of data science, impacting diverse sectors from healthcare to societal governance. Our approach, rooted in collaborative design, sought to ensure theoretical soundness and practical applicability in varied data science contexts. This process involved a structured formation of a working group, in-depth literature reviews, brainstorming, and expert consultations, enriching TAPS-RM with diverse perspectives and practical insights. This document aims to transparently portray the TAPS-RM's evolutionary process, highlighting its initial conception, iterative refinements through expert feedback, and practical validation. Serving as an informative guide, TAPS-RM seeks to help researchers, practitioners, and policymakers in understanding and applying TAPS-RM's principles to promote responsible data science practices.

**2. Preliminary key terms and concepts**

Our research work into TAPS-RM is part of a broader initiative undertaken at the Institute of Data Science (IDS). At IDS, we undertake several research projects with social/legal and ethical impacts. This realization led us to identify an important gap: How do we define the responsibility components in our data science projects? We work on technological innovations and research questions such as: How can individuals be given greater control over their data? How to analyze data in a privacy-preserving way? These research questions also require that we explore how varied responsibility agendas influence innovation. For instance, how might privacy-preserving technologies differ across regions and countries with different ethical and legal approaches?

Since its founding in 2017, IDS has pursued research questions that fall under “responsible data science by design,” such as advancing the FAIR principles. In 2019, our TAPS-RM group was funded to further formalize how a responsibility agenda should not be “bolted on” to technological innovation but threaded throughout a data science project.

The aim of the group was to define: What is responsible data science by design?

**Responsible data science emerges when transparency, privacy/confidentiality, accountability, and social values are implemented and evaluated across:**

* actors (people and organizations)
* objects (data, algorithms, knowledge graphs, protocols)
* data science processes (procuring, analyzing, and disseminating data), and
* impacts (the intended and unintended consequences in the world).

***"By design" refers to the need for:***

* ongoing ethical reflection
* stakeholder and organizational commitments
* responsive technology development, and
* evaluative mechanisms.

**3. Matrix development stages**

**Stage 1: Working group formation and Summary of initial brainstorming sessions, focusing on key ideas, debates, and consensus reached.**

* *Working group formation*

The initial step in developing TAPS-RM was to set up a small working group (10 people). The selection of the members was based on their knowledge in the domain of data science and its wide-ranging applications across numerous sectors such as ethics, law and society, health, life sciences, agrifood systems, sustainability, environment, and data management. The selection procedure entails a detailed review of their present and past contributions in their respective fields, an assessment of the members' participation in relevant data science projects, and recommendation from their peers.

* *Initial steps of the working group*

The formation of the working group marked a pivotal step in developing the TAPS-RM. The kick of meeting of the working group centering on the concept of ‘responsible by design’ in data science projects. The initial meetings focused on the theme of practicing ‘responsible by design’ in data science projects and exploring the prerequisites for responsible data science. Discussions covered various aspects of responsibility, such as biases, impacts, and the need for transparency in data science methods and analysis. Key discussion points included:

* Recognizing the impacts of data/ training data on societal influences, which can sometimes lead to unnoticed biases.
* Responsibility involves controlling and being aware of biases, not only in the training dataset but also in variable selection and model assessment.
* Emphasizing the importance of understanding a tool’s impact to prevent mismatches between intended goals and potential side effects.
* Stressing the necessity of making robust statements that exhibit long-term responsibility.
* Raising questions about data ownership/control and how these factors impact decision-making processes.
* Examining the relevant social unit of responsibility (individual, family, community) and the supporting political philosophies and cultural norms

Following these initial and foundational discussions, the working group started by identifying the potential dimensions and scopes that encompass responsible approaches in data science. We described various “components” representing a spectrum of principles in data science and “scopes” that encapsulate existing and emerging practices in the domain. This step was crucial in uniformly covering the expanse of responsible data science by design, setting the stage for the subsequent development of the TAPS-RM.

**Stage 2: Literature review and refinement of concepts**

In the second stage of our research, we conducted an extensive literature review to deepen our understanding of the current state of responsible data science practices across various domains. This crucial step involved identifying established practices and confirming the foundational elements that enhance the value of methodologies in responsible data science. Our approach to the literature review was methodical and thorough. We initiated our search with a set of carefully chosen keywords and phrases, specifically tailored to our area of focus within responsible data science. To ensure that we captured a broad spectrum of relevant literature, we explored multiple academic databases, covering publications from 2014 to 2020. The selection process entailed a preliminary screening of titles and abstracts for relevance, followed by an in-depth examination of the full text to determine the appropriateness of each publication for our study. We gave priority to works that explicitly discussed aspects of responsibility in data science or that provided tangible guidelines for adopting responsible practices.

From this process, we analyzed 28 seed publications, spanning from 2014 to 2020. This analysis provided a concise summary of their perspectives on various facets of responsibility and suggested best practices for responsible data science. Notably, our review also highlighted three prominent responsibility frameworks in the field: FACT, FAIR, and Data sheets for Datasets. These frameworks played a significant role in guiding the evaluation of our responsibility matrix later stage. Furthermore, this literature review solidifies our comprehension of four principal components of the responsibility matrix. We established preliminary definitions for each of these components, selecting and refining them for their direct relevance to responsible data science practices. Supported by insights gleaned from the literature, we further refined the definitions of these four components as follows:

**Transparency:** Transparency plays an important role in the reproducibility of research results. It allows other researchers to check the validity of scientific methods, to ensure that desired results can be achieved and to prevent misinterpretation of the results. The concept of transparency in science is to communicate the necessary details about a study clearly, openly and accurately. Transparency in data science would refer to openly and clearly specifying and documenting the research purposes and objectives for the data science study and also openly describing their methodology to achieve these objectives.

**Privacy/confidentiality:** Modern data societies are evolving the sense that certain uses of data might challenge the autonomy of data subjects, and there are several existing concepts for talking about that autonomy.

1. **Privacy** has been defined from various points of view such as law, organization, and philosophy. However, there is no universally accepted definition of privacy. To give a relatively generic description of privacy as: i) the right of a person to determine which personal information about himself/herself may be communicated to others: ii)the control over access to information about oneself, iii) the freedom from intrusion into one's personal matters and personal information.
2. **Confidentiality** refers to personal and non-personal information shared with a select group or for an explicit purpose that generally cannot be divulged to third parties without the expressed consent of the individual data subject. This consent is granted with the expectation that it will not be revealed to others or used for non-compatible purposes.

While confidentiality is an ethical duty, privacy is a right rooted in the common law. For example, the General Data Protection Regulation 2016/679 (GDPR) is a regulation in the EU law on data protection and privacy in the European Union and the European Economic Area. GDPR is a mechanism for satisfying these ideas about the autonomy of the data subject.

**Social values** Societies operate according to ethical norms and social values that are also reflected in responsible approaches to technology broadly and data science in particular. These ethical norms and social values can involve the preservation of human life, the importance of innovation and technological progress, the cohesion of communities, the improvement of health, the sovereignty of individuals and their freedom to choose, and so forth. Societies differ in which norms and values they prioritize, and societies often hold inconsistent or paradoxical norms and values. These norms and values also change, often in unpredictable ways.

As far as data science goes, projects are often undertaken to support and respect those values. The pursuit of these norms and values are often used to motivate data science projects, and they are used to justify trade-offs between interests that may be at odds with each other.

These definitions were further refined and completed with additional work. We refer to the paper as the up to date version of our findings.

**Accountability** Accountability pertains to how a particular responsibility agenda is implemented and empowered within the structure of an organization. An organization will show its commitments to supporting responsible data science in its internal ethical stances and disciplinary processes, its budgetary priorities, and organization of teams.

There are a range of ways for organizations to implement a responsibility agenda. If roles for actors within data science projects have been established, and if they know how to execute those roles sufficiently, then they can be equipped with necessary knowledge and skills to apply preventative and corrective measures for malpractice in the particular part of a data science project that they lead or conduct. They can also be made aware of what potential social, legal and ethical pitfalls each process in the data science pipeline is prone to. There would be a clear social, legal and ethical code to which each actor in a project must abide.

An organization could establish measures for analyzing the degree to which actors and organizations abide by their responsibilities. It could also require educational programmes to prevent accidental malpractice as well as disciplinary procedures for deliberate and persistent malpractice.

Ultimately, accountability is based on the seriousness with which a society takes transparency, privacy, confidentiality, and the preservation of its values in data-related projects. It will also be based on the extent to which societies want data to be useful.

**Stage 3: Expert consultation and feedback integration**

Following the literature review stage, qualitative research with the aim to verify and fine-tune our approach was performed. Experts were selected by identifying either projects that included a data science component or by identifying experts known to contribute to the data science field with multiple projects. The experts shared trades such as secured research funding, impactful publication records, they were performing high leadership roles and participating in advisory and committee roles.

Prior to the interview, we shared the definitions of the components and scopes with the selected experts via email, as they were formulated in Stage 2. We asked the experts to review these definitions in preparation for the interview. At the beginning of the interview, a brief explanation of how we defined the responsibility agenda for data science: the components of responsibility (transparency, privacy/confidentiality, accountability, social values) and their application to the scopes (humans and organizations, digital objects, data science processes, social impact) were explained to the expert.

Semi-structured interviews were individually performed with every selected expert and two interviewees. The second interviewee was added to remove bias and consolidate interpretation of the results. The semi-structured contained several questions driving the conversation:

1. How significant do you believe transparency (encompassing privacy and confidentiality), accountability, and social values are for data scientists? Additionally, which of these components do you regard as the most crucial, and why?
2. Reflecting on your extensive experience with various data science projects in your domain, what were your initial thoughts on different key components, and how, if at all, did your perspective on these components evolve by the end of these projects?
3. During your data science projects, have you encountered any conflicts with collaborators from various domains concerning various components?
4. Does the shared definition of components encompass all aspects that you consider crucial? Additionally, how would you define this component in terms of responsibility within your field?
5. Considering your experiences in data science projects and your broader professional journey are there any insights you have acknowledged that you believe should be included in the definitions of the four key components? How do you see these insights fitting into the existing framework of these components?"

All the participants found that the components and their descriptions were relevant and thematically coherent, but they provided feedback about limitations of the initial draft of the TAPS-RM. For instance, we learned the importance of addressing liability in the “accountability" components. The interviews gave us confidence that the four components capture all important aspects of responsible data science by design.

We did not explicitly test the scope definitions with the experts. However, feedback was analyzed to refine the component definitions and think about the components in relation to the scopes: For example - When asked about the transparency component, subject 2 stated: *“In medicine it’s important to explain the limits (Where can it go wrong?). Be open about how it works. Specify the intended use. Deep learning is a big example where you can’t explain things.* This brought us to the reflection “Where in the TAPS-RM matrix we capture this information?” In TAPS-RM, the scopes were named Agents, Objects, Processes and Results. We already were thinking of results as a way to define wanted and unwanted outcomes. This meant that the naming of the scope “results” had a narrower meaning than our intentions. In our follow up meetings, we further discussed this and determined that the scope should be named “Impacts” to reflect how outcomes are intended to be used but also how they might be mis-used.

**5. Testing and refinement process**

Once we had the initial TAPS-RM components and scopes refined, we used an iterative process to test and further refine the matrix. Multiple methods were alternated and used in this process during the three years of the TAPS-RM development. For example, presentations of the work to other groups and continuous testing to seek further feedback and to improve the work. The following summarizes the work in a stepwise process[[1]](#footnote-1):

**Step 1 - Group Testing**: We tested the matrix with selected projects from the members of the group (n 6). The involved members of the group filled in the matrix for a specific project that they were working on at the time. Feedback was provided in the group meetings to further improve the matrix, especially the understanding of the intersection between the components and the scopes. During this testing step, we further clarified the definitions of the components, how to separate a data science project in terms of actors, objects, processes and outcomes and consequently what information belongs to each cell.

**Step 2 - Larger group testing.** Following our first experiment a workshop was organized to present the TAPS-RM matrix to our own data science institute (n.20), After the presentation of the matrix, we asked the audience to fill in the matrix for their own projects. The feedback from the session was that there was a need to make the matrix more accessible to different data science projects. In order to guide and trigger the response of the users in describing the various components for their projects, we decided that including further questions in all the cells of the TAPS-RM matric was the best way to evolve the framework. Several discussions arose during this step. On one hand, leaving the dimensions open was a good way to be generic and to leave the users focus on what was most relevant in their domain. On the other hand, prompting questions would limit the scope to specific questions. that could better capture and define the responsibilities. We decided that questions had various benefits such as standardizing the TAPS-RM framework and making comparisons across project responsibilities more comparable. For each cell in the matrix, we created a first version of the questions. In several of the follow up meetings of the group we discussed and revised the questions that were included in TAPS-RM.

**Step 3 - User Questionnaire.** A Questionnaire was developed to check how well users understand the questions and which components scopes are more relevant to the users to identify if there is a specific order they would approach the Matrix (n.48). The questionnaire is still available on [this link](https://maastrichtuniversity.eu.qualtrics.com/jfe/form/SV_eLju4fLyLmbEqWy). The results indicated that while all components were important, transparency was broadly indicated as the most important component to be reflected on first and the most relevant scope was “Impacts”. Furthermore, the questions that were unclear for the users were identified. These questions were later further explored in a user study.

**Step 4 - Focus Group.** To further test and define the questions of TAPS-RM, a focus group meeting was organized (n.5) with data scientists working on projects in different domains (health data analysis, economics and law-tech). The identified participants were selected outside from our own research institute. We wanted to better understand how to address unclear questions as well as if there is a natural order to answering the questions of the TAPS-RM matrix. The findings enabled us to further refine the questions of the matrix, however the responses on the best order of the questions remained mixed. This led us to conclude that the user interface of TAPS-RM should be flexible and enable users to jump to the cells and to questions as needed.

**Step 5 - Development and testing of the user interface:** After step 3 and 4, we could develop the TAPS-RM matrix as a [web-based user interface](https://responsability-matrix.137.120.31.102.nip.io) showing the questions in the matrix and enabling users to answer them for their project. We further tested the matrix within workshops where the participants (n.16) filled in the questionnaire using existing EU or National ongoing projects from different domains (healthcare, law, economics, and agri-food sector). The feedback was used to further improve the questions of TAPS-RM.

**6. Mapping with existing initiatives**

A reflective summary of the journey from the inception to the finalization of the TAPS-RM.

We mapped questions/statements of FACT, FAIR and Data sheets for datasets to the high level cells of TAPS-RM (i.e Given a statement from FAIR, which Component/Scope does it belong to?).

We chose to evaluate the matrix with these three frameworks because they ranged from concisely formalized (FAIR) to expansive (Datasheets for Datasets), and because their creators ranged from international consortia (FAIR, Datasheets for Datasets) to single authors (FACT). More specifically, the FAIR guidelines were selected because they provide increasingly important guidance on an international level to how researchers produce digital research objects. Experiences of our group in their dissemination and implementation of FAIR also inform the formulation of the concept of the responsibility matrix. The FACT guidelines were chosen because they have been prominently circulated in research circles. The “Datasheets for Datasets” guidelines were an intriguingly comprehensive approach that mirrored the responsibility matrix in its use of open questions that researchers would ask themselves. The principles, guidelines, and recommendations from the three frameworks were located in the cells of the TAPS-RM matrix. The matrix format shows where these frameworks overlapped with each other and where gaps exist.

The mapping for each framework was made independently by teams of two, if the mapping did not match then it was discussed in the larger group. Two researchers independently mapped the chosen guidelines into the matrix. Then, they compared the answers and identified discrepancies. The goal was not to resolve these discrepancies in judgment but rather to identify them and note them as places where either 1) the responsibility matrix was ambiguous or 2) the specific guidelines were ambiguous.

The details of the mapping and the findings are highlighted in the paper.

1. Some of these steps might not appear explicitly in the paper as in the paper we selected the most relevant steps to describe the work. [↑](#footnote-ref-1)