Paper DH06

GDPR: Risks and Recommendations from the PHUSE Data Transparency Working Group

Shannon Labout, Data Science Solutions LLC, USA
Arlene Coleman, Pfizer, USA

ABSTRACT

PHUSE convened a project in 2018 to research and discuss the impact on clinical research of the General Data Protection Regulation (GDPR) and to produce a white paper that would provide recommendations on applying GDPR rules to collection of Personally Identifiable Information (PII) and to safeguards and processes for the protection of clinical research data. This paper has been prepared by the GDPR project sub-team leads and will discuss our research on how organizations are implementing safeguards into their processes and how they are approaching collection of personal or sensitive data. The draft recommendations that will be in the PHUSE GDPR White Paper will be presented includes recommendations for what data should be collected, who should have access to that data during and after the study and a discussion of the risks, with various data collection modalities.

INTRODUCTION

The General Data Protections Regulation (GDPR) from the European Parliament was published in 2016 and became enforceable in 2018. GDPR replaces the European Data Protection Directive 95/46/EC and its purpose is to assure the privacy and security of personal data collected in the process of any business operation, including clinical research. GDPR includes several provisions that strengthen consent requirements and grant significant new rights to citizens of European countries, including the right to be notified about data breaches, right to access their data, and the right to erase their data (often called "the right to be forgotten"). On the surface these new rights seem to pose significant burdens on the clinical research community, so the interpretation and application of the regulation in clinical research has been a hot topic of conversation for the past couple of years.

PHUSE convened a Project in 2018 to research and discuss this and to produce a white paper that will provide some guidance on applying GDPR rules in clinical research. This paper will introduce the findings of the GDPR Project and describe the process and results of the research behind the white paper.

One important thing to note is that GDPR includes some exemptions for special categories of data, of which clinical research data is one. This does not mean that clinical research data would be exempt from all provisions of GDPR, but certain provisions (such as the right to erasure) would not apply. A discussion of those exemptions is outside the scope of this paper, as are some related (but important) processes, such as informed consent.

PHUSE GDPR PROJECT

The PHUSE Data Transparency Working Group has worked on several projects, including developing a Data De-Identification Standard for SDTM and creating a Data Transparency Roadmap. In 2018 the Data Transparency Working Group started a project to address the impact of GDPR on clinical research data collection practices. The GDPR Project was further divided into 3 sub-teams for the purpose of sharing experiences, approaches, recommendations and understanding of GDPR compliance related to:

DATA COLLECTION AND PII

The purpose of this sub team is to recommend and share practices on data collection activities for GDPR alignment and compliance, including how to handle the collection of personally identifying information (PII) such as birth date or age across age ranges and study types, and what not to collect or to collect differently in any study. Under GDPR "PII" is referred to as "personal data". Personal data is the focus of GDPR, so discussions about data collection in this paper are limited to personal data and do not include other data, including anonymized data. GDPR.org describes personal data as *"…any information relating to an identifiable person who can be directly or indirectly identified."*

SAFEGUARD AND PROCESSES

This sub team focused on topics including GDPR requirement implications, data sharing and data breach recommendations, long-term approach of archiving and storing data, other country regulations specific to GDPR, and overall process, documentation and communication specific to protecting patient data privacy. Collaborations across industry have shared approaches and best practices as the sub-team continues to summarize recommendations specific to safeguarding patient level data.

DATA BREACH

This sub team initially met to focus on data breaches, database access and communication issues and how the sharing of data works between clinical sites and CRO's. They quickly realized there was significant overlap with the topics being discussed and researched by the Safeguard and Processes sub team, so the Data Breach sub team disbanded and merged with the other sub teams.

After the initial research, discussions and drafts, the GDPR Project sub teams worked together to document GDPR best practices for clinical research in the form of a white paper that is currently in public review and is targeted for publication by the end of 2019.

DATA COLLECTION TOPICS

One of the primary goals of GDPR is to empower individuals to have more control over their own data, and much of that aspect relies on a robust informed consent process, which is outside the scope of this paper. Although clinical research is a special category of data, and some of the GDPR provisions are limited for clinical research, such as the right to erasure of data, the spirit of GDPR should be considered throughout the lifecycle of clinical data. The main recommendations from the Data Collection and PII team are related to putting appropriate limits on how much personal data should be collected.

The Data Collection and PII sub team focused on developing working definitions for terms that would be used in the White Paper, identifying commonly collected data points that could potentially be PII, and discussing the handling of these data points. As part of the process, this team conducted a survey on current industry practices around the collection of birth dates and age. The results of that survey are discussed in this paper and will be detailed in the PHUSE GDPR White Paper.

RESEARCH - DATA COLLECTION

The Data Collection and PII team researched the impact of GDPR on data collection activities by reading the GDPR documentation, sharing information about their respective organizations' approach to GDPR, and through informal public surveys shared through LinkedIn and by email. The main focus areas for this team were:

1. Identifying information that would be considered personal data with the potential to allow tracing data back to an individual research participant
2. Researching current practices and recommendations for minimizing the collection of personal data and mitigating the risks associated with collecting that data.
3. Identifying opportunities for transforming data collection practices in keeping with the "spirit" of GDPR.

In pharmaceutical, biotechnology and medical device research, the research team is typically masked from having any personal data about the research participants, evidenced by the long-standing practice of using masked identifiers for participants rather than using their actual names. Furthermore, over the past 20 years, data collection practices have evolved along with variable privacy laws to minimize the risks of being able to identify an actual person from collected data. To illustrate that point, consider the way we identified research participants in the 1990s - typically with their full birth date and their actual initials. Over the past couple of decades these practices have evolved to the point that it is now more common to collect only a year of birth, and "subject initials" are not typically collected by most pharmaceutical, biotechnology or medical device manufacturers or their research partners due to privacy concerns and interpretation of predicate regulations related to privacy (e.g., European Data Protection Directive 95/46/EC).

Because the collection of personal data had already become quite restrictive before GDPR, the Data Collection team were interested to learn how, or if, GDPR would actually change data collection practices. For most organizations, the only personal data being collected is the Birth Date, so the primary question was how organizations would collect Age or Birth Date under GDPR regulations, and what special rules would they apply for pediatric studies. The companies

represented on the Data Collection and PII team contributed to this conversation, but we wanted broader input, so the team conducted a public survey with the following questions:

1. Under GDPR, do you plan to continue collecting any components of Birth Date in any of your data, including EDC, lab requisitions, ePRO/eCOA, etc?
2. If you collect only a sub-component of Birth Date, how do you complete it?
    a. Month not completed at all
    b. Month completed with a random number
    c. Month completed with a fixed number (specify)
    d. Day not completed at all
    e. Day completed with a random number
    f. Day completed with a fixed number (specify)
    g. Other
3. Do you plan to collect Age (e.g., 6 Weeks or 46 Years) instead of Birth Date?
4. Do you plan to collect Age in addition to Birth Date?
5. What is the maximum age at which you would consider it appropriate to collect a complete Birth Date (Year, Month and Day)?
6. What is the maximum age at which you would consider it appropriate to collect both Month and Year of Birth?
7. Is the maximum age for which you would collect Birth Month based on your local regulations and definition of when someone becomes an adult and can sign their own informed consent?
8. At what age does a person become an adult in your local region?

Responses were received from a total of 30 respondents representing at least 12 different pharmaceutical, biotechnology and CRO companies. 3 respondents refused to disclose their company name, but one indicated that they were employed by a top 10 global pharma company. Responses came from at least USA, Europe and Japan.

The results of that survey revealed the following:

- 69% plan to continue collecting some components of actual Birth Date
- 36% will collect only Birth Year, another 36% will collect Birth Month and Year
- 57% will collect Age (either in addition to or instead of Birth Date)

Most respondents skipped some or all of the rest of the survey questions

- Out of 13 respondents, there was no consensus on the age at which a complete Birth Date could be collected, but ~46% would not collect complete Birth Date past the age of 12, and another 23% would not collect complete Birth Date past the age of 18. Respondents indicated that the collection of complete Birth Date for pediatric studies would depend on the needs of the study
- Out of 22 respondents, the responses were fairly evenly split on the rationale for setting these age limits on the collection of complete Birth Date. About half said that the age limit was based on local regulations or definitions of “adult”, and the other half indicated other reasons for the age limit.
- Out of 22 respondents, 20 responded that “18” is the age at which a person becomes an adult in their region.

RECOMMENDATIONS - DATA COLLECTION

Limit the collection of data to what is necessary. Multiple studies have demonstrated that it does not take much data to identify an individual from a large group of people. Having a birth date, sex and a relatively close postal code is usually enough. Since clinical studies may require the collection of sensitive personal data GDPR requires us to consider carefully what data points are actually needed to support the requirements of the research study, and to obtain clear “opt in” consent from participants before collecting their data.

As previously mentioned, clinical research is a special category of data collection and usage, so some of the GDPR rules (e.g., the right to erasure / ‘right to be forgotten’) do not apply. However, that does not mean we should continue to collect data when we are not using it. To meet the ‘data minimization’ principle of GDPR, we should limit

the collection of clinical data to that which is necessary to performing an analysis for the study, assessing the safety of the drug, or other valid research purposes that have been disclosed to the participant through informed consent. For example, if ethnic origin will not be used to explore the treatment in ethnically-based patient populations, and there is no reason to believe certain ethnic groups would have a different safety profile, ethnic origin should not be collected.

Follow predicate rules for data integrity and provenance. GDPR does not overrule current regulations and rules, such as 21 CFR Part 11, so, for example, we should not actually hard delete data from a database if a participant withdraws consent from the study or withdraws consent from further data collection. GDPR does not replace 21 CFR Part 11; however, it does put stronger requirements in place for ensuring that we do not collect unnecessary data in the first place, for informing the participant of the intended uses of their data, and for giving them access to the data they have consented to provide.

Identify the risks for your studies and have a mitigation plan. Understanding the risks that are specific to your research is an important step to understanding how to apply GDPR in your organization. For example, if you are using any data collection methodology that allows free text entries (which most organizations do), one potential risk is that someone will enter an actual name or phone number of a research participant. Having documentation (e.g., SOP) for what to do in that situation, and training the appropriate people on that SOP shows that you are planning to comply with regulations. A table is provided in the white paper with potential risks and suggested mitigation plans for each identified risk.

SAFEGUARD AND PROCESSES TOPICS

The Safeguard and Processes team focused on topics surrounding safeguarding patient study data to ensure GDPR compliance. The team conducted a survey to obtain valuable input from PHUSE members and cross industry participants. Survey results were summarized and recommendations included in the GDPR White Paper.

RESEARCH - SAFEGUARD AND PROCESS

The Safeguard and Processes team wanted to understand GDPR's impact on how we handle data for its entire lifecycle. The team followed a similar method of research by reading GDPR and related documentation, sharing experience from their organizations and by conducting an informal survey online. The questions in that survey were:

2. Who should have access to view files containing subjects' personal data during the conduct of the study?
    a. In general, sponsor and CRO staff who are part of the study team
    b. In general, sponsor and CRO staff who work on studies for the relevant molecule
    c. In general, sponsor and CRO staff who work in the relevant therapeutic area
    d. In general, sponsor and CRO staff in biometrics functions
3. Who should have access to view files containing subjects' personal data after the end of the study?
    a. In general, sponsor and CRO staff who are part of the study team
    b. In general, sponsor and CRO staff who work on studies for the relevant molecule
    c. In general, sponsor and CRO staff who work in the relevant therapeutic area
    d. In general, sponsor and CRO staff in biometrics functions
4. Are any further restrictions appropriate?
    a. Allow access only to a subset of the data for the study depending on the user's role
    b. Requirement to renew access on a periodic basis
5. After the end/submission of the study, is it appropriate to retain files containing personal data outside the Trial Master File (TMF) for any of the following reasons:
    a. As a backup in case gaps are found in the TMF
    b. To provide additional details to internal staff or regulatory agencies if needed
    c. To save resource burden reviewing files that are already in a restricted-access area
6. If any such files are retained after the end/submission of the study, how long would be an appropriate length of time?
7. What should happen to the study datasets and the other TMF files that contain personal data at the end of the retention period required by the relevant regulations?
    a. Destroyed
    b. Anonymized
    c. Retained

8. What measures should be taken by a sponsor to document and communicate the safeguarding of data by a CRO on their behalf?
    a. Agreement or contract
    b. Training
    c. Manual of operations
    d. IT security assessment
9. What are appropriate and inappropriate practices in discussing subject data and data issues? (For example, is it appropriate for certain data discussions between the sponsor and CRO to take place via email?)
10. If anonymized study data is to be shared with external researchers, what should the process involve in order to meet data protection requirements?
11. What data breach scenarios are plausible in the context of clinical trials, and what preventative measures are appropriate for these?
12. What training, processes/SOPs or plans should be put in place to prepare to respond in the case of a data breach?
13. Are there any aspects of the collection, management or storage of trial data not mentioned above that might be affected by GDPR?
14. Are there areas in which it is appropriate to apply a different approach to GDPR implementation to past versus ongoing versus future studies?

A total of 23 respondents from at least 12 different pharmaceutical, CRO and vendor organizations participated in the survey. Respondents represented multiple regions including USA, UK, EU, Japan, China and India.

Many of the survey questions were open-ended, so the team reviewed and summarized those responses and included the consensus responses in the recommendations. Some of the structured responses revealed a strong consensus, e.g.,

- For Question 1: Who should have access to view files containing subjects' personal data during the conduct of the study? 81% responded that in general, access to study files should be limited to sponsor and CRO staff who are part of the study team
- For Question 3: Are any further restrictions appropriate? 74% responded that access should be to a subset of data based on user role, and 68% said access should be reviewed and renewed periodically. Some added further requirements to restrict all access to anonymized data only
- For Question 4: After the end/submission of the study, is it appropriate to retain files containing personal data outside the Trial Master File (TMF) for any of the following reasons: 80% responded that providing additional details to internal staff or regulatory agencies was an appropriate reason to retain files after the end of the study/submission.
- For Question 6: What should happen to the study datasets and the other TMF files that contain personal data at the end of the retention period required by the relevant regulations? 80% responded that anonymizing data was preferred so that the data could serve secondary purposes.

Some of the questions with structured responses had more nuanced results, e.g.,

- For Question 2: Who should have access to view files containing subjects' personal data after the end of the study? While 57% responded that In general, access should still be restricted to the sponsor and CRO staff who are part of the study team, 28.5% added that access could be expanded to include sponsor and CRO staff who work on studies for the relevant molecule or therapeutic area, and another 9.5% thought access could be expanded to include any sponsor and CRO staff in biometrics functions.
- For Question 7: What measures should be taken by a sponsor to document and communicate the safeguarding of data by a CRO on their behalf? 95.5% responded with Agreement or contract, 64% with Training, 68% with a Manual of operations, and 77% with IT security assessment indicating that all of these methods are important.

RECOMMENDATIONS - SAFEGUARD AND PROCESS

The primary recommendations coming from the Safeguard and Processes team are related to access restrictions and retention of personal data.

Access Restrictions:

This team recommends that access to files that contain personal data should be controlled by both technology and process and should generally be limited to the study team members at the Sponsor and CRO. Other staff can be approved to have access, but outside the study team access should be granted only for qualified research activities and then ideally only to appropriate subsets of anonymized data.

Access should be specified and controlled by contractual agreement, IT security assessments, Manual of Operations and through training. The combination of these approaches should ensure that data are protected from intentional or accidental breach throughout its lifecycle, so

- Technical system access controls should be employed
- Process access controls should be in place, and all relevant staff trained
- High risk methods of communication, such as email, should not be allowed for exchanging files with personal data.
- Formal contractual agreements should clearly specify each party's responsibilities with respect to personal data protection

Retention of Personal Data:
The first step is always obtaining informed consent from the participant, making it clear to that person whether the organization intends to retain their personal data beyond the end of the study, and the intended or potential uses for their personal data beyond the end of the study.

Data retention length should be based on regulatory requirements and business needs, and ideally the data will be anonymized for any secondary (or beyond) uses.

CONCLUSION

GDPR should change the way we think about data collection and the protection of data for its entire life cycle. We should acknowledge that without the people who participate in research who willingly agree to allow the use of their data for research we would not be able to conduct research, and we should willingly and intentionally respect their contributions by protecting their privacy.

Limiting data collection to what is necessary for the analysis is the most important step that an organization can take to protect a research participant's privacy. If we do not collect personal data, we cannot lose it or accidentally expose it. GDPR requires us to think about what we actually need to collect and then to limit data collection to what is necessary for the analysis. We should then do everything we can to safeguard the data we collect through technical and process controls, and by having a clear plan and process for retiring (destroying or anonymizing) the data when we no longer need it.

REFERENCES

GDPR Websites:
https://eugdpr.org/

https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en

PHUSE Data Transparency Working Group Wiki:
https://www.phusewiki.org/wiki/index.php?title=Data_Transparency

ACKNOWLEDGMENTS

The authors would like to acknowledge the entire PHUSE Data Transparency team for their support during the process of developing, reviewing, revising and publishing this White Paper, and especially the dedication and contributions of the GDPR Project team.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:
Author Name: Shannon Labout

Company: Data Science Solutions LLC
Address: 223 Salt Lick Road, Suite 279
City / Postcode: St. Peters, MO 63376
Work Phone: 917-821-7165
Email: Shannon.Labout@DataScienceSolutionsLLC.com
Web: www.DataScienceSolutionsLLC.com

Author Name: Arlene Coleman
Company: Pfizer, Inc.
Address: New York, USA
Work Phone: 862-324-1036
Email: Arlene.coleman@pfizer.com
Web: www.pfizer.com