Recommendations for GDPR Compliancy

Contents

Overview: Purpose of this document ........................................ 1
Scope ..................................................................................... 1
Definitions ............................................................................. 1
Problem Statement .............................................................. 1
Background ............................................................................ 1
Recommendations ............................................................... 2
Disclaimer .............................................................................. 9
Appendices and Sources ..................................................... 9
Acknowledgments ............................................................... 9
Overview: Purpose of this document

The General Data Protection Regulation (GDPR) from the European Parliament was published in 2016 and became enforceable in 2018, replacing the 1995 Data Protection Directive (95/46/EC). The GDPR’s purpose is to assure the privacy and security of personal data collected in the process of any business operation.

In 2018 the PHUSE Data Transparency Working Group convened a project to investigate how clinical research organizations are interpreting and applying the GDPR rules within their organizational structures, technologies and processes. This paper has been prepared jointly by the GDPR Project Sub-teams to offer practical recommendations for applying the GDPR rules to the collection and management of clinical study data throughout its lifecycle. The recommendations are based on the experience of sub-team members and the results of two online surveys conducted in 2018 to obtain a wider range of opinions from within the industry.

Scope

This paper covers the practical application of the GDPR to the collection, processing and storage of clinical study participant research data by clinical study sponsors and their outsourcing partners, including secondary data use. Specific information around the legal basis for processing research data, and the application of the GDPR to source data retained at sites and to the data of staff working on studies are not within the scope of the paper.

Definitions

The following are definitions of personal and sensitive data that were part of the previous (now superseded) regulation Directive 95/46/EC. The bolded language has been added as part of the GDPR updates:

• Personal Data: Any information relating to an identified or identifiable natural person (’data participant’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.

NOTE: This is different from ‘Personally Identifying Information’, a term used in North America, which is more limited in scope.

• Special Categories of Data (’Sensitive Data’ under the previous directive):
  • Personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade-union membership;
  • genetic data;
  • biometric data [e.g., facial images] if processed for uniquely identifying a natural person;
  • data concerning health;
  • data concerning a natural person’s sex life or sexual orientation.

In addition to the above updates, the EU has also provided clarification (bolded) to how it defines anonymous and pseudonymous data in the GDPR:

• Anonymous Data: Information which does not relate to an identified or identifiable natural person, or personal data rendered anonymous in such a manner that the data participant is no longer identifiable.

The GDPR does not concern the processing of anonymous or anonymized data once it has been rendered anonymized.

• Pseudonymous Data: Personal data that can no longer be attributed to a specific data participant without the use of additional information, provided that such additional information is kept separately and is subject to technical and organizational measures to ensure that the personal data is not attributed to an identified or identifiable natural person.

NOTE: This data is still considered personal data, but the GDPR recognizes that the application of pseudonymization can reduce the risks to the data participants concerned and help controllers and processors to meet their data-protection obligations.

NOTE: In the context of clinical trials and based on recommendations in this white paper, even when relatively specific identifiers such as participant initials are not collected, study data should usually be regarded as pseudonymous, not anonymous, because of the presence of study participant IDs (which are linked to the participants’ names at the clinical site) and/or quasi identifiers (e.g., a combination of variables that together could be used to identify an individual).

Problem Statement

Recommend appropriate processes to support GDPR compliance in the collection and subsequent handling of clinical study data.

Background

In 1995, the European Union (EU) adopted the Data Protection Directive (Directive 95/46/EC) as a means by which the privacy and protection of all personal data collected for and about its citizens would be protected. However, in 1995, the internet was still in its early stages. The General Data Protection Regulation (GDPR) is an evolution of Directive 95/46/EC that was formally adopted by the EU on April 27th, 2016 and became enforceable on May 25th, 2018.

The purpose of the GDPR is to harmonize EU member states in their approach to data privacy. Each member state may impose additional requirements or obligations in addition to the GDPR. The EU also created a Data Protection Board for enforcement of the new regulations, as well as a complaint process for EU residents. Data breaches now must be reported to the
appropriate authority within 72 hours of identification and fines can be up to 4% of a company's global revenue (or 20m euros, whichever is greater).

The objective of the GDPR and associated supporting functions is to increase individual protections, which include allowing greater transparency to EU residents, along with reporting requirements to regulators. The GDPR allows EU residents to have more control over their personal data.

The rules in the GDPR better meet the needs of the digital world by addressing how data is collected, stored and transferred. It provides citizens with more control over their personal data by giving them two new protections – “right to be forgotten” and “right of portability” (i.e., opt-in/opt-out clauses) – and requiring clear consent and justification for personal data usage.

The GDPR applies to the personal data (information) of persons in the EU and EEA, wherever the data is collected or processed.

In the clinical study context, the processing of personal data must comply not only with the GDPR but also with other data protection and clinical study regulations. The GDPR specifically states that processing of personal data for scientific purposes should comply with other relevant legislation such as legislation on clinical trials. It also states that individual member states (countries) may produce their own guidance for research and clinical study regulations. The GDPR specifically states that processing of personal data for scientific purposes must comply not only with the GDPR but also with other data protection regulations for the clinical study context. See CNIL’s MR-001 and MR-003.

Recommendations

Data Collection Recommendations

To meet the intent of the GDPR, the following high-level recommendations are offered:

1. **Follow all current laws that govern data integrity and provenance.** Continue to follow current regulations, such as U.S. 21 CFR Part 11, which are intended to ensure that data is protected from unauthorized access or use, that the systems used to collect, store and transfer data work as they are intended, and that there is a clear history of the data. The GDPR does not replace these regulations; it establishes strong rules for ensuring that we do not collect unnecessary data and that we properly handle any data we do collect.

2. **Limit collection of data to necessary data only.** To meet the ‘data minimization’ principle of the 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 purposes. The principle of ‘data minimization’ is a principle that the GDPR particularly highlights as important in the context of scientific research. France’s MR-001 and MR-003 list the specific variables that may be collected in a clinical study context.

Multiple studies have shown that it does not take much data to identify an individual from thousands or millions of other people. Latanya Sweeney writes “About half of the U.S. population (132 million of 248 million or 53%) are likely to be uniquely identified by only (place, gender, birth date), where place is basically the city, town, or municipality in which the person resides. And even at the county level, (county, gender, birth date) are likely to uniquely identify 18% of the U.S. population. In general, few characteristics are needed to uniquely identify a person.” Given that we know this, it should be a matter of course that we strive to collect only the data we actually need to perform analysis on clinical research results. In clinical research much of the data collected also falls under the GDPR definition of Special Categories of Personal Data, which require additional protection.

3. **Ensure relevant documents are clear (e.g., protocols, informed consent forms (ICFs), electronic data capture (EDC) forms) that a participant’s decision to stop study drug, study participation and/or data collection are separate decisions.** Organizations should clearly define the difference between stopping study treatment early, stopping the study early and withdrawing consent for data collection or processing in the protocol. Under the GDPR, withdrawal of consent for data collection or processing refers to a participant’s right to no longer have data, including biological samples, collected, processed or stored, which is different from a participant’s desire to end their participation in a clinical study (e.g., undergoing medical procedures for study purposes) or a subject’s desire to stop taking the study drug. Consideration should be given to no longer using the standard terminology “Subject withdrew consent” in the CRF, and instead using more specific language to indicate from “what” the participant subject is withdrawing consent (e.g., “Subject withdrew consent from study participation”, “Subject withdrew consent from data collection and processing”).

These high-level recommendations apply to any type, method or source of clinical research data collection. Systems that are specifically created for clinical research data collection (e.g., “EDC”) usually must meet compliance with the current regulations that govern all electronic systems in clinical research. Any data collection system may have additional risks to consider and should be evaluated for potential risks.
Recommendations for GDPR Compliancy

Table 1 provides examples of potential risks which should be addressed through standard operating procedures (SOPs), training and documentation (e.g., Data Management Plan), along with recommendations for risk mitigation.

Table 1: Data Collection Risks and Mitigation Plans

<table>
<thead>
<tr>
<th>Risk</th>
<th>Mitigation Plan</th>
</tr>
</thead>
<tbody>
<tr>
<td>User may enter directly identifying data (e.g., participant name)</td>
<td>Cover this risk with an SOP and training for site staff and other individuals who may provide data (e.g., central readers, adjudicators, labs and other vendors). The SOP may include the following steps for mitigation:</td>
</tr>
<tr>
<td>(e.g., participant name) into a free-text field (e.g., eDiary, EDC,</td>
<td>If the data is identified before database lock:</td>
</tr>
<tr>
<td>Serious Adverse Event narrative) or erroneously send a file</td>
<td>• Redact the data to minimize the risk of further individuals seeing it. (In EDC, this may mean prompting the site to remove it from the front end of the database via a query. If the direct identifier was in the attachment to an email, it may involve deleting the email and asking for it to be re-sent.)</td>
</tr>
<tr>
<td>containing such data (e.g., admission/discharge summaries, death</td>
<td>• Evaluate with a knowledgeable attorney and document how the record of such data entry in any clinical study database audit trails will be handled according to applicable predicate rules and regulations (e.g., in the United States, 21 CFR Part 11 would take precedence, but in the EU, the GDPR may take precedence in some circumstances and require it to be redacted with a statement).</td>
</tr>
<tr>
<td>certificates for event adjudication) to the sponsor or CRO (contract</td>
<td>• Immediately notify and re-train the person who entered the data through normal study processes (e.g., Contract Research Associate (CRA) re-trains site staff).</td>
</tr>
<tr>
<td>resource organization)</td>
<td>• Notify the organization’s Data Protection Officer (DPO) and document the incident.</td>
</tr>
<tr>
<td>Data transfers from some vendors (e.g., wearables, ePROs) may include</td>
<td>Ask the vendor to limit the amount of information required to be entered in the wearable to that which is necessary to produce analysis results.</td>
</tr>
<tr>
<td>full name, address, email, phone number, GPS location information or</td>
<td>Ensure that directly identifying information is redacted from data transfers so that none of this information is transferred to the sponsor or CRO.</td>
</tr>
<tr>
<td>other direct identifiers</td>
<td></td>
</tr>
<tr>
<td>Lab samples may require participant initials, full birth date and</td>
<td>Work with lab vendors to update their requisition forms to require only necessary information to process the samples. There are some cases (e.g., young pediatrics) in which a complete birth date or sex would be necessary to apply reference ranges precisely, but in most cases year of birth (and sex) should be enough. If full birth date does not need to be collected, it may be possible to remove this from the data before the lab transfers the data to the sponsor or CRO.</td>
</tr>
<tr>
<td>sex</td>
<td></td>
</tr>
<tr>
<td>Source documents (e.g., admission/discharge summaries, death</td>
<td>Ensure this information is redacted in these documents if they need to be transmitted outside the clinical site by having an appropriate SOP and providing training to sites, CRAs and other staff involved in handling source documents.</td>
</tr>
<tr>
<td>certificates for event adjudication) often include direct identifiers</td>
<td></td>
</tr>
<tr>
<td>It may be difficult to control data collected from certain medical</td>
<td>Work with data providers to modify the machine settings so that the data can be collected in a pseudonymized way if possible, maintaining traceability to the individual research participant while minimizing the use of direct or indirect identifiers.</td>
</tr>
<tr>
<td>devices (e.g., ECGs, MRIs) that are designed for use in the</td>
<td>If the machine settings cannot be modified to require no directly identifying information, the data transfer to the sponsor or CRO should have this information redacted.</td>
</tr>
<tr>
<td>participant setting. These devices may be set up to require minimum</td>
<td></td>
</tr>
<tr>
<td>information (e.g., birth date, initials) or direct identifiers (name,</td>
<td></td>
</tr>
<tr>
<td>individual’s tax identifier, address) in order to take measurement</td>
<td></td>
</tr>
<tr>
<td>or make the assessment</td>
<td></td>
</tr>
<tr>
<td>Rare disease populations are a unique challenge because the number of</td>
<td>Implementing the high-level recommendations and all the risk mitigation strategies described in this table will help.</td>
</tr>
<tr>
<td>participants is small</td>
<td></td>
</tr>
</tbody>
</table>

Since the question of how precise a subject’s age should be for sufficient data integrity may depend on the age of the study population, there are additional considerations that should be made. The following are recommendations for what to consider when deciding how much information to collect for age:

1. Do a risk assessment – how you collect the data (e.g., if you include day and month in your EDC form because it is allowed to collect those in some clinical sites, you run the risk of having sites in countries with stricter privacy rules inadvertently violating those rules by entering data).
2. In conversations with your medical and clinical staff, biostatisticians and other stakeholders, consider what you need to collect for the study endpoints and safety, and collect only that much and no more.
3. A survey conducted by PHUSE to inform the content of this white paper indicated that most companies are interpreting privacy rules as meaning that only age or only year of birth (possibly both) should be collected, nothing more (except in special cases, such as pediatric studies).
4. Pediatrics: Base your approach to collecting birth date for pediatrics on
   a. Previously mentioned general principles (what data is required)
b. Age group (may need very specific age for lab ranges in very young infants/neonates)

c. Therapeutic area (e.g., growth studies may need more specific age than month and year. ICH requirement to track pediatric participants’ weight and height against standard growth charts, need their age in number of months)

d. For most pediatrics (if there are no other factors that would require more data):
   i. Up to 2 years old, collect full birth date
   ii. From 2 years to 18 years, collect month and year

e. Pediatrics is typically considered to be:
   i. Up to 18 years (but may be different in some countries)
   ii. Other factors: other tests to determine when growth plates are mature.

5. Recommendations for the precision of birth data collection may be different for different data sources (e.g., EDC vs. safety labs).

6. Recommendations for reconciling data across data sources:
   a. Lab, ECG, EDC, randomization data
      i. Lab requisitions may not be flexible – may require full birth date. Some organizations impute a month and day in their EDC to perform age calculations (but will have to drop the imputation before populating standardized (i.e., Study Data Tabulation Model “SDTM”) submission datasets).
      ii. ECG machines may require full birth date.
      iii. EDC and Interactive Response Technologies (IRT) can be set up to collect year only (or year and month).
   b. Make risk-based decisions around when/what to query (e.g., more than a year discrepancy) based on study.
   c. Consider adding recommendations for checking exact age (e.g., to ensure the eligibility of participants near the age threshold) via source documents, as part of a monitoring plan.

Safeguarding and Processes Recommendations

Access Restrictions

Access to files that contain participants’ personal data should generally be limited to the designated study team members including the sponsor and/or CRO and/or the sponsor’s designee(s). If additional staff members outside of the clinical study team require access to the clinical data, this can be provided once reviewed and approved by appropriate designated approvers. This can include individuals who may work in that study’s therapeutic area or similar target molecule for decision-making and other required analysis. Informed consent and primary/secondary analysis definitions within the study protocol require review to ensure participants have agreed to how their data can be used, including for any purposes beyond the original study such as pooled analyses. Anonymization of data is required if data may be shared more widely, depending on the research activities. Good Clinical Practices should continue to be followed.

It is suggested to allow access to only a subset of data (if applicable to the needs) and renew clinical study data access on a periodic basis.

Further restrictions to clinical study data access may apply on a case-by-case basis to ensure compliance when needing access for pooling of study information, exploratory research and secondary use considerations. As mentioned, the appropriate approval process outlined should be followed prior to obtaining access to clinical study data.

Retention of Personal Data

Retention of any files that may contain personal data should be in accordance with the policies set forth by the sponsor and should always respect the commitment made to the participant in the informed consent form.

Sponsors should consider the purpose of retaining any files containing personal data beyond the retention periods required by clinical study regulations. It may be necessary to retain certain non-TMF files generated during the study for a reasonable time after the study ends to ensure or improve quality of documentation or processes in the future. Below are a few recommendations:

• The purpose of such retention should be documented.
• The retention should be covered by the sponsor’s retention policies.
• Ensure the consent from the participant covered the length and purpose of the retention.
• The sponsor should only retain minimum data that is required.
• Periodic review of the retained data, as per the review schedule of the sponsor, is necessary.

At the end of the retention period, sponsors should re-evaluate the policies that apply to the retained data and consider the privacy risk, scientific value and clinical relevance of the data for future research. We live in a data-oriented world where past clinical trial data may be useful in discovering cures/new solutions for chronic and rare diseases. If deemed useful by the sponsor for future research of such diseases, it may be possible for sponsors to retain study data if it is anonymized.

Sponsor-led Documentation to Ensure Safeguarding of Data is Communicated to a CRO

Four options were presented to the survey respondents, along with an opportunity to enter free-text measures. The options presented were the agreement or contract, training, manual of operations, and IT security assessment.

The agreement or contract was almost unanimously agreed an essential method of safeguarding data. The agreement introduces data safeguarding obligations in many forms e.g., legal actionability, indemnity, relationship, or operational dependence. The contract is also the foundation upon which the sponsor/CRO relationship is built.

An IT security assessment was also considered very important when sharing data with a CRO. Sponsors expect CROs to have robust network security procedures, firewalls, issue resolution protocols, and an auditable record of data accessors.

A Manual of Operations (MoP) was less frequently regarded as a method of communicating the safeguarding of data. This can potentially be attributed to the somewhat broad scope of MoPs, and other documents that may cover data safeguarding e.g., risk management plans, SOPs, etc.

Training is highly recommended for data safeguarding
Recommendations for GDPR Compliance

awareness, with the following considerations:

**CRO-led activity:** Training may not be created or driven by the sponsor. CROs and other large organizations generally have extensive intra-company training programs to ensure Good Clinical Practice (GCP) and other compliance. Furthermore, internal CRO policy and culture may drive strict training requirements that meet the sponsor’s needs. The sponsor may consider the CRO’s record of successful training completion among employees sufficient for data safeguarding, while other companies require additional company-specific training.

**Time and cost:** Refresher or additional training is not typically suggested as a first option due to associated time and costs but may be mandatory as country-specific regulations continue to evolve and change. Robust documentation, trackability, and contractual obligations are typically completed as risk-management or preventative measures; training may be a downstream step if the other methods are found insufficient. Training is commonly associated with a Corrective and Preventive Action (CAPA) process that arises when an audit or review exposes an issue. CAPAs are followed closely by health authorities, requiring considerable time and effort, which is why avoidance of an issue/finding is key.

Training under a different name: Standard operating procedures may be attributed to basic onboarding, or distribution and review of company procedures. There may also be trust in educational and professional certifications that can obviate the need for training in more basic practice but may still be a company-specific requirement to complete additional training.

Other important measures that were volunteered in the free-text response box include:

- Sponsor review of CRO’s training requirements and SOPs
- Clean desk policy: This is indicative of a CRO’s company culture of data hygiene and prevention of unexpected leaks
- Audit of IT physical plant – security in physical hot storage, cold storage, and archive environments must be reviewed in addition to cybersecurity measures
- Storage and archive of final documents and reports
- Periodic checks to ensure policies are being followed
- Permanent data storage at sponsor; data deletion from CRO.

**Practices in Discussing Participant Clinical Data and Clinical Data Issues**

The survey responses indicated a range of opinions on possible practices in these areas, with varying levels of associated risk to the security and compliance of the participant information. Practices range from permissive (allowing more free discussion/communication of the participant information) and restrictive (regulating or prohibiting similar communication):

**1. Data discussions over email**

Historically, this has been raised as a risk for several reasons, including:

- Risk of inadvertently sharing by errors or unexpected changes to mailing lists;
- Emails can be forwarded and widely distributed without restricted access;
- Emails can be subpoenaed for legal proceedings, risking exposure of participant-level data that may not have been required for the legal proceeding, but is now part of public record;
- Cybercrime interception risk (or the need to encrypt to mitigate this).
Range of responses from survey respondents:

### Data Discussions over Email

<table>
<thead>
<tr>
<th>Email should be OK</th>
<th>Secure encrypted and traceable email that meets HIPAA &amp; GDPR requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes, as long as a particular patient’s identifying data is not then put into the email</td>
<td>Subject IDs should never be included in email</td>
</tr>
<tr>
<td>Limit identification to subject ID only</td>
<td>No</td>
</tr>
</tbody>
</table>

#### Permissive

- Appropriately, so long as the data exchanged cannot be followed back to the actual person

#### Restrictive

- Depends on how identifiable the individual may be
- There should be no discussion concerning subject data (especially if indentifiable) over email
- Limit identification to subject ID only

### 2. Access controls and examples of acceptable practices

Nearly all responders were in favor of access controls, though there is variation over what content can be shared rather permissively versus what requires access control. Verbal discussions are commonplace and accepted practice, as stated by multiple survey responders. It is beneficial to address discussions in a meeting room. Tighter access was recommended for written or recorded material.

Several themes emerged across recommended practices – there was high consensus on establishing formal data management and access process or procedures to minimize sharing of or unauthorized access to personal data. Access controls and training elements were also suggested as data security practices.

Range of responses from survey respondents:

### Access Controls and Acceptable Practices

<table>
<thead>
<tr>
<th>Data cannot be re-identifiable. A question regarding visit x for subject y is unlikely to contain enough information to be identifiable</th>
<th>Subject data should be shared and discussed only through secure means, using a minimal amount of information necessary to identify the subject under discussion (i.e., subject ID)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOPs are in place to remove any inadvertent violations of PHI in email</td>
<td>Documents to be filed with the source data</td>
</tr>
<tr>
<td>CRO or vendor must be trained in data security, and ensure only the appropriate parties gain access to the material shared</td>
<td>Have scheduled calls to discuss subject data and publish minutes</td>
</tr>
</tbody>
</table>

#### Permissive

- Study data (eg, assessment outcomes) are not specifically PHI
- There should be protocols in place for sharing personal data via email
- Need to avoid risk of unauthorized access

#### Restrictive

- Only generalized discussion should be part of an email and not have specific questions related to the data
- Communication should flow through a communications application that meets the security and tracking standards set forth by regulatory requirements and security guidelines
- Data management process required
- Read-only data sharing of source data with access controls
Recommendations for GDPR Compliancy

3. Recommended file transfer modalities

Several file transfer modalities were mentioned by respondents to accommodate the need for data security and discretion in recipient and viewer. These include:

- a. secure data transfer mechanism
- b. password-protected files
- c. recipient validation
- d. encryption
- e. read-only data sharing of source data with access controls.

It should be noted that in certain legal scenarios (e.g., litigations) data may need to be shared beyond the original intended recipients of the files.

Data breach scenarios, preventative measures and how to prepare

Where possible, ensure that study data shared or transferred does not include certain pieces of information that could be used to trace back to an individual (e.g., birth date, absolute age, postal code, rare events, etc.). IT security controls should be in place to reduce this risk, along with legal contracts to assist with deliberate and accidental breaches. Security control examples include setting up firewalls for internal systems, transferring data in a secure fashion, multi-factor authentication, and using cybersecurity best practices in providing access.

There are also new cybersecurity companies that monitor data flow, create white label IP, and implant honeypots (a computer security mechanism) to detect and distract attempts at unauthorized use of information systems.

An effective preventative measure is providing GDPR-specific training for personnel and CRAs to ensure standard operating procedures are being followed, and how to address accidental breaches such as CAPA and preventative/corrective measures for GDPR audit readiness. CAPA specific measures could include root cause, assessment of damage and impact to each stakeholder, communications plan, execution to completion of corrective measures, broad IT/business collaboration for appropriate preventative measures, project management to ensure timely follow-through of prevention, wrap-up and lessons learned.

Verify periodically that personnel are up to date with training, test IT systems for security, ensure systems have robust data audit trails, conduct awareness training on a regular basis, conduct workshops to make the situation ‘real’ to all employees, make SOPs user-friendly, and avoid recording/maintaining/sharing personal data (unless data is anonymized and approved to share). An organization should have a standing data security/data privacy policy in place that outlines the steps to limit and safeguard data.

These preventative measures can cover potential data breaches such as:

- cyber-attacks affecting electronic data capture systems
- sharing/release of laboratory/testing record
- exposure of printed documents or computer screens
- inexperienced sites sending identifying information in an inappropriate fashion
- ungoverned sharing of participant data between sponsors, CROs and third parties
- mobile devices storing participant-level data being lost or stolen
- malicious individuals posting participant-level clinical trial data in a publicly available site
- data being accessed by an unauthorized individual and transferred to an outside party
- researchers re-identifying participants
- life-threatening conditions of a participant increasing the likelihood of identification of a participant by their acquaintances.

If a data breach occurs:

Ensure the data breach is brought immediately to the attention of the relevant staff by following appropriate SOPs and procedures specific to data breach activities. A detailed Standard Operating Procedure (SOP) documenting the data breach procedure can help ensure the relevant parties responsible for receiving, investigating and reporting data breaches are clear on their roles and responsibilities. If there are ever any concerns, consider the role of a DPO to empower the organization in this context.

Additional aspects of the collection, management or storage of trial data not mentioned above that might be affected by the GDPR

- The ICF should be updated with new GDPR requirements and informed consent language.
- Determine and document what happens if a participant wants to exercise any of their privacy rights.
- Study staff and site investigator data is also personal data. Addresses, contractual arrangements, and CVs can be in multiple departments/locations for employees at an organization.
- Using cloud-based globally accessible servers vs using traditional localized servers.
- Conduct Privacy Impact Assessments to assess the risk of clinical trial data processing in your company.
- Novel data processing activities such as artificial intelligence and application of machine learning using previously collected real data.
- Creation and governance of large-scale data repositories for access and analysis, e.g., “Data Lake”.

Sharing Anonymized Study Data with External Researchers

Review and approval of external data sharing

Data requested to be shared with an external party should be approved by a governing body prior to sharing. An impartial board should assess research proposals and data requests. This can be likened to an Institutional Review Board or Ethics Board – a multidisciplinary expert panel that evaluates the request based upon several criteria. At a minimum, the assessment should include whether the data request is for an unmet need, whether the data requestors are qualified to perform the planned analysis, and whether participant consent has been given for that research purpose. To evaluate in an unbiased manner, researchers wishing to use the data should document their data access request with several elements:
Recommendations for GDPR Compliancy

- research project goals
- specific data requested (including data not from trial)
- procedure/statistical analysis plan and/or project endpoints
- data elements or variables essential to the project
- list of named accessors of the data, and proof of their ability to conduct analysis (e.g., CV or publication list)

Patient consent

The first contract that has been created on this data is that with the participant. Participant consent on data use should not be violated when participant clinical data is shared with external researchers. Affiliates, subsidiaries, and sub-contractors that require data access toward conduct of a trial are typically covered under participant consent. However, reuse of the participant data by researchers toward aims not specified in the trial plan may not be covered under the consent form. Patient informed consent forms (ICFs) should be reviewed and compared to the scientific workplan of the researchers.

A factor to consider is that interpretations differ between countries, each state within the US, and regulatory bodies in the interpretation of informed consent forms toward data reuse. A comparison of the research plan should be conducted against the study’s ICFs by country or US state where possible. Further fragmentation between entities is expected in the years beyond GDPR regulations, especially as non-EU countries adopt similar legislation.

GDPR and data minimization principle

The data minimization principle states that personal data shall be “adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed.” Applied to sharing with external researchers, only clinical datasets that are needed to reach the listed scientific aims should be shared. Only provide the required subset of variables and datatypes specific for the request prior to the anonymization step. A review of the scientific plan against the data requested can potentially reduce the amount of data unnecessarily shared. Qualifications of data accessors can also justify that the parties receiving the data to breach, especially if left unguarded in a transient environment. Once data has reached the destination, it should immediately be removed from the transient location (e.g., Box, secure FTP, hard drive).

Access controls (e.g., password protection, approval process, two-step authentication) should be employed when accessing datasets. This is described in the area on data sharing via email. Data access should be limited to the intended dataset for the planned analysis; it should not allow the user to view all data available in the storage or analysis area.

An auditable trail to identify how the data was transferred is recommended, along with the use of GxP tools and environments for data transfer and analysis. GxP environments will also enable auditing the data for accessors and can be compared to analysis intent.

Contractual elements and access requirements

The sponsor should always have a contract dictating the terms and conditions of data sharing and access. Master Service Agreements and Material Transfer Agreements may cover these elements, especially for service provision or subcontractors using data. For academic or other secondary researchers, safeguards around data access and purpose should be clearly agreed in writing.

Further details can be captured in a Data Sharing Agreement. Items covered in this agreement may include:
- modality of data access and/or transfer
- named accessors of the data
- length of access term
- prohibitions on sharing the data with parties not listed
- safeguards around researchers discussing and sharing data with each other
- sponsors retaining the right of refusal for manuscripts, or further ethics reviews (if applicable) for data privacy reasons only
- protocols in the event of unauthorized access/use, or accidental or intentional data breaches
- data deletion/destruction requirements, and proof of data destruction
- rights to further use of the data
- prevention of data recipient attempts to re-identify the data
- prohibiting attempts to re-identify a participant, or development of an AI/machine learning algorithm that enables re-identification of a participant.

It is also important to note that a Data Sharing Agreement should be employed when data is shared with peer-reviewed journals.
De-identification/anonymization process

Data shared outside the company/sponsor for purposes outside trial conduct is safest when anonymized. The first step of anonymization is to generate a new alphanumeric or scannable code for the participant data and destroying the link between the two. Beyond this, companies should keep internal business rules toward anonymization of participant data. Please refer to Appendices for additional reference information.

Quantitative risk assessment for external sharing of data is an industry trend in response to the GDPR and increased participant privacy requirements. This involves the calculation of a risk threshold for re-identification of the participant, and de-identifying the data to a likelihood below the calculated threshold. Tools and methods for such de-identification are being developed as companies adopt this practice for data sharing.

When double coding identifiers, identifiable variables should be removed and/or banded. Common items to remove include absolute dates, initials, state identifiers, location, rare disease status, or rare events. There is legislation in place guiding which variables cannot be shared, and several recommendations have been published. Data controllers should redact direct identifiers and manage indirect identifiers so equivalent class sizes are appropriate.

For example, age and BMI can be banded into ranges. Locations can be generalized to global region e.g., Asia-Pacific, Americas, Africa-Middle East, Europe. It is best to understand the variables necessary for the planned analysis so they can be retained as much as possible in the anonymization process. Datasets shared should also remove sensitive participant data that is easily identifiable. Rare events or rare disease phenotype participants can be removed from the dataset shared, if not required for analysis.

Multiple linked data types add another dimension to the de-identification process. Data types may include clinical data from a trial, genetic data, image/scan data, video data, or data collected in real time from mobile devices and wearable items (e.g., step counters, blood glucose monitors). These should be considered holistically upon de-identification so that a link remains between the various data types to be shared. Applying different approaches to GDPR implementation for past, ongoing and future studies

Applying different approaches to GDPR implementation for past, ongoing and future studies

For future studies, only collect what is necessary to limit exposure. Generally, the data from past studies should be treated according to the same privacy standards as future studies, though clearly it is not possible to provide updated privacy notices to participants for past/closed studies. Identify unrequired files containing personal data to ensure data protection governance is completed for past and ongoing studies.

Obligations, regulations, ethics committee requirements for individual studies may differ. Appropriate review and approval from legal and data stewards within a company should be completed along with appropriate procedures and documentation.

Conclusion

The wide scope and general nature of the principles of the GDPR leave room for differences in its practical application in the clinical study context. However, the overall intent is clear: To respect individuals, especially as technological advances change the way data can be used. The PHUSE GDPR project hopes that this paper will help organizations consider how best to handle the data of clinical study participants to ultimately contribute to the furtherance of human health and wellbeing.

Disclaimer

“This is an interpretation of the GDPR with suggestions for possible ways to apply it in practice based on opinions from individuals in the PHUSE project and unidentified survey respondents. It does not constitute legal advice or recommendations from individual companies.”

Appendices and Sources

PHUSE Data Transparency Working Group Wiki

GDPR Website

EU Data Protection Rules

History of the GDPR

De-identification and Anonymization of Individual Patient Data in Clinical Studies - A Model Approach
https://www.phuse.eu/white-papers

End Notes


Acknowledgments:

Authors: Arlene Coleman, Shannon Labout, Hannah Sharp, Ashwini Weber, Margi Sheth, Michelle Brooks

PHUSE Data Transparency Working Group

Brand and product names are trademarks of their respective companies.