CASE STUDY

Data protection record of processing activities and privacy notice generator toolkit by EMBL’s European Bioinformatics Institute [version 1; peer review: awaiting peer review]

Montserrat González Ferreiro, Steven Newhouse

Technical Services Cluster, EMBL European Bioinformatics Institute, Hinxton, Cambridgeshire, UK

First published: 05 May 2022, 11:500
Latest published: 05 May 2022, 11:500

Abstract

EMBL-EBI, Europe's biomolecular data hub, is a world leader in managing and analysing big data in biology and making it freely available to scientists worldwide. Researchers can access the open data resources and related services of EMBL-EBI by submitting minimal personal data. In May 2018, following the enforcement of the European Data Protection Regulation (GDPR), EMBL adopted the EMBL Internal Policy no. 68 on General Data Protection. It reflects European data protection principles while remaining within the bounds of EMBL's international legal status. As a result of GDPR and EMBL's Internal Policy No. 68 coming into force, 190 EMBL-EBI user-facing services that processed personal data in 2018 were required to have Records of Processing Activities (RoPA) and Privacy Notices (PN). EMBL-EBI's solution was to develop a Data Protection Engine (DPE) that automatically generates RoPA and PN when a service owner answers a series of questions. In addition to maintaining a centrally located database for RoPAs and PNs, the DPE tracks changes to the documents, as well as providing versioning and time-stamped updates. It is the aim of this article to share the EMBL-EBI IT department's experience with designing and implementing the DPE and providing a toolkit to let others develop a similar solution and benefit from our experience. Implementation steps, benefits, challenges, opportunities, and practices are discussed and critically analysed.

Keywords

Data protection, personal data, GDPR, research infrastructures, intergovernmental organisations, privacy notice, record of processing activities.
Corresponding author: Montserrat González Ferreiro (montserratgferreiro@gmail.com)

Author roles: González Ferreiro M: Conceptualization, Formal Analysis, Methodology, Project Administration, Resources, Software, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; Newhouse S: Formal Analysis, Supervision, Validation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2022 González Ferreiro M and Newhouse S. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: González Ferreiro M and Newhouse S. Data protection record of processing activities and privacy notice generator toolkit by EMBL’s European Bioinformatics Institute [version 1; peer review: awaiting peer review] F1000Research 2022, 11:500 https://doi.org/10.12688/f1000research.121363.1

First published: 05 May 2022, 11:500 https://doi.org/10.12688/f1000research.121363.1
**Introduction**

The European Molecular Biology Laboratory (EMBL) is the leading intergovernmental laboratory for life science research in Europe. Aimed at advancing the study and understanding of molecular biology, training young scientists and fostering innovation in science and technology alike. It has six sites in five host nations - one of which is EMBL’s European Bioinformatics Institute (EMBL-EBI).

EMBL is an intergovernmental organisation, subject to international public law, entrusted with a number of privileges and immunities necessary for its functions. Accordingly, it has the power to self-regulate data protection. Such self-regulation is necessary to take account of EMBL’s status as an intergovernmental organisation, and its focus on scientific research beyond national borders.¹ The EMBL Internal Policy no. 68 on General Data Protection defines EMBL’s self-regulation in response to the requirements of the European General Data Protection Regulation (GDPR), which came into force in May 2018.²

As a result of GDPR and the coming EMBL Internal Policy no. 68 requirements, EMBL-EBI had to prepare nearly 200 Records of Processing Activities (RoPA) and Privacy Notices (PN) for data processing activities, including scientific and IT services, within less than four months. To expedite and automate this process, EMBL-EBI developed the Data Protection Engine (DPE), which uses a set of questions and templates to automatically generate all the required documents.

**Steps to design the Data Protection Engine**

To simplify the design of the DPE, we used a form to collect the necessary information from Service Owners regarding their data processing activities in order to generate RoPAs and PNs automatically. The information collected in the form is used to complete templates for RoPAs and PNs.

With the templates and questionnaire ready, we defined the DPE workflow to facilitate collaboration between staff, service owners, and data protection specialists.

As soon as the workflow design was complete, the implementation began, and notifications were added to improve workflow management.

**Creating the templates for Records of Processing Activities**

The RoPAs are inventories of the personal data that is being processed and how it is being processed. In our analysis, we determined that we needed at least two RoPA templates: one for EMBL-EBI to use as a data controller, the other for EMBL-EBI to use as a data processor. On occasion, EMBL-EBI acts as a joint data controller, but these cases are not as common. In such cases, EMBL-EBI will share the data controller RoPA template with other controllers so they can use it to jointly define the content and shape.

The template for EMBL-EBI as a data controller reflects that it is processing the personal data and determines the purposes and means of the processing of personal data (see Template 1, Extended data ³). There are three main sections in the template: data controller information, data processor information, and data protection impact assessment.

The template for EMBL-EBI as a data processor reflects that it is processing the personal data on behalf of a data controller, and the purposes and means of the processing of personal data defined by the data controller (see Template 2, Extended data ³). There are two main sections in the template: data processor information, and data protection impact assessment.

The information required to fill out the RoPA templates has been identified, and different scenarios have been considered. The options included in the templates also reflect some of the most common scenarios, such as the types of personal data processed and their purposes.

**Creating the templates for Privacy Notices**

The PN offers data subjects (in our case service users) information about how an organisation processes personal data and complies with data protection principles. Two PN templates were created. The first PN template is for use by EMBL-EBI when it is a data controller, and it collects personal data directly from data subjects. The second template is to be used when EMBL-EBI is a data controller and it collects personal data indirectly, such as when another organisation provides the data. When EMBL-EBI performs the function of a joint data controller, it shares the PN template with other controllers in order to define content and structure jointly.
The PN template for a data controller who obtains personal data directly from data subjects (see Template 3, Extended data) and the PN template for a data controller who does not obtain personal data directly from data subjects (see Template 4, Extended data) have the same structure. Both include seven sections: who controls the personal data and how to contact them, the lawful basis for processing personal data, how personal data is collected and used, who has access to the personal data, data transfers to third parties or international organisations, data retention, and data subject rights.

Different cases have been analysed to determine what information is needed to fill out the PN templates. The options included in the templates also reflect some of the most common scenarios, like who will have access to the personal data.

Summary of the documents to be generated by the DPE
In different scenarios, different results will be generated by the DPE. Mainly, we have to consider where the personal data comes from in addition to the freedom to decide how the data is to be processed. Documents generated by the DPE are summarised in Table 1.

In the case of EMBL-EBI being a Joint data controller with other organisations, both the RoPA and the PN will be created using the DPE manual mode, a mode developed to create tailored documents.

Creating the Data Protection Engine questionnaire
The questionnaire includes all questions necessary to obtain the information that can be used to complete the templates, as well as some information that is useful for monitoring purposes or to communicate with an appropriate contact.

Because the DPE was created when the GDPR was to come into force and there was a high level of uncertainty regarding the exact requirements, we wanted a comprehensive questionnaire to allow us to expand the RoPAs and PNs as needed.

The questionnaire (see Template 5, Extended data) includes nine main sections that include data on the data processor or data controller, basic information on the service that processes the personal data, sources of personal data, personal data processing responsibilities, time limits for erasure of personal data, technical and security measures used to protect personal data, a short data protection impact assessment, and the lawful basis for processing personal data.

The questions are written considering EMBL and EMBL-EBI, however they can be easily adapted to request information about the processing of personal data by other organisations, as can the templates.

Designing the EMBL-EBI Data Protection Engine workflow
The key aim of the DPE is to facilitate the creation and review of RoPAs and PNs and to coordinate the interactions of the many roles involved: service owner, team leader and data protection administrator (DP Admin). The service owner is the operational contact for the service. A service owner reports to the team leader, who is ultimately responsible for the service processing personal data. Data protection administrators provide support to the service owner and team lead on all data protection questions and review the information they provide in the DPE.

| Table 1. Outcomes produced by the Data Protection Engine (DPE), mainly Privacy Notices (PN) and Records of Processing Activities (RoPA) based on the personal data origin and level of freedom for European Molecular Biology Laboratory-European Bioinformatics Institute to decide how to process it. |
|--------------------------------------------------|--------------------------------------------------|--------------------------------------------------|
| **EMBL-EBI decides how to process personal data** | **Personal data obtained from users** | **Personal data not obtained from users (e.g. from other organisations)** |
| RoPA as a Data Controller PN for data obtained from data subjects | RoPA as a Data Controller PN for data obtained from data subjects |
| **EMBL-EBI has to agree with other organisations how to process the personal data** | RoPA as a Joint Data Controller PN for data obtained from data subjects (they will be generated using the DPE manual mode) | RoPA as a Joint Data Controller PN for data not obtained from data subjects |
| **EMBL-EBI processes the personal data based on instructions from other organisations** | RoPA as a Data Processor PN for data obtained from data subjects | RoPA as Data Processor |
The following workflow shows how to use the DPE, from reading the user guidelines to publishing the RoPA and PN with the automatic mode, when the questionnaire generates the RoPA, and when the content needs to be customised (Figure 1).

The process for manual and automatic creation is nearly identical, with only one difference: the second step for manual creation already requires the involvement of the data protection specialist, who holds the role of DP Admin, and reviews all responses to the questionnaire, called the DPE record. When the manual mode is selected, there is a warning displayed to the users saying “Manual mode will freeze all the answers to the questionnaire not being able to be modified later and keep all the selected values. Only the Data Protection administrator can edit the Record of Processing Activities and Privacy Notice manually”.

Manual creation also has the disadvantage that any changes to the RoPA and PN will have to be made manually. Automatic creation, on the other hand, only requires updating relevant fields in the questionnaire to generate new versions of these documents.

There are notifications going to the Service Owner, Team Leader, and Data Protection Administrators, as we can see from the workflow. To ensure that there is always someone available to assist DPE users, the notifications to the Data Protection Administrators can include a list of people.

Data Protection Engine current status and future work

In January 2022, we had 259 Records of Processing Activities and Privacy Notices, of which 10 were customised. Overall, the DPE has proved its usefulness with the vast majority of data processing activities following common patterns. Our decision to provide a manual option for the few uncommon cases that cannot be covered with the automatic approach appears justified.

In the near future, we intend to improve notifications and send reminders to the data protection support team if requests aren’t handled after one day. Additionally, we would like to make the options provided by the questionnaire regarding personal data retention periods more specific.

It will also be necessary to review the templates with the EMBL Data Protection Officer in the long term, since the requirements of EMBL IP 68 and GDPR are clearer today.

In the future, the DPE could be expanded as a data protection service hub by designating Data Protection Assessment as a separate entity from RoPA and PN, including links to documentation and training for EMBL staff, on demand video

---

**Figure 1.** Data Protection Engine (DPE) Workflow to create automatic or tailored Records of Processing Activities (RoPA) and Privacy Notices (PN). The workflow involves actions by Data Protection Administrator (DP Admin), Service Owner and Group Team Leader (GTL) among others. This figure is an original figure produced by the authors for this review article.
training for new joiners, and gathering together all the guidelines and instructions for implementation that we have collected over the years.

**Conclusion**

The toolkit includes the questionnaire to collect information from service owners and store it in a central repository, the templates to generate RoPA and PN, a recommendation to offer the possibility to create manual RoPAs and PN if they need to be tailored, for example if a specific format or look and feel is required, and the workflow designed to develop the DPE. This project saved EMBL-EBI considerable time and helped standardise data protection practices, and we are confident that it can help other organisations as well.

**Data availability**

**Underlying data**

No underlying data are associated with this article.

**Extended data**

Open Science Framework: Supplementary materials for the Data Protection Record of Processing Activities and Privacy Notice generator toolkit by EMBL’s European Bioinformatics Institute. [https://doi.org/10.17605/OSF.IO/S856G](https://doi.org/10.17605/OSF.IO/S856G).

This project contains the following extended data:

- 01-DP-Article-Template1-RoPA-Data Controller.docx
- 02-DP-Article-Template2-RoPA-DataProcessor.docx
- 03-DP-Article-Template3-PN-DataFromDataSubject.docx
- 04-DP-Article-Template4-PN-DataNotFromDataSubject.docx
- 05-DP-Article-Questionnaire.docx

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

**Acknowledgements**

Joseph Rossetto and Liang Shen, along with other team members of the EMBL-EBI Web Development team, contributed intellectually to the development of the EMBL-EBI's Data Protection Engine in 2018. We would also like to thank Daniel Gant for reviewing this article's grammar. An earlier version of this article can be found on OSF preprints (DOI: 10.31219/osf.io/3wbez).

**References**

The benefits of publishing with F1000Research:

• Your article is published within days, with no editorial bias
• You can publish traditional articles, null/negative results, case reports, data notes and more
• The peer review process is transparent and collaborative
• Your article is indexed in PubMed after passing peer review
• Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com