**Draft: 2.6 October 2018**

**SCOR Global Life**

Principles for ensuring compliance with legal obligations on the Experience Analysis Platform

Document Information

|  |  |  |
| --- | --- | --- |
| ***Properties*** | The document properties are: | |
|  | Owner | Karsten de Braaf |
|  | In force from | [•] |
|  | Author | Karsten de Braaf and Heather McKay |
|  | Approved by | SGL CRO & SGL Chief Legal Counsel (TBC) |

|  |  |
| --- | --- |
| ***Review Process*** | This paper shall be reviewed at least every three years.  Unless they are minor, changes made to this paper are subject to prior approval by the Experience Analysis Central Review Team.  The table below outlines the changes included in the latest three versions of this paper (including the present version). |

|  |  |  |
| --- | --- | --- |
| **Version & date** | **Approver** | **Main changes** |
|  |  |  |
|  |  |  |

|  |  |
| --- | --- |
| ***Scope*** | All users of the Experience Analysis Platform |

Table of Contents

[Introduction to the Experience Analysis Platform 4](#_Toc528256865)

[Legal Considerations 9](#_Toc528256866)

[Anonymisation Guidelines 10](#_Toc528256867)

[Annex A – Data dictionary 15](#_Toc528256868)

[Annex B – Standard Rule 15](#_Toc528256869)

[Test 1: UK 15](#_Toc528256870)

[Test 2: France 16](#_Toc528256871)

[Test 3: Germany 17](#_Toc528256872)

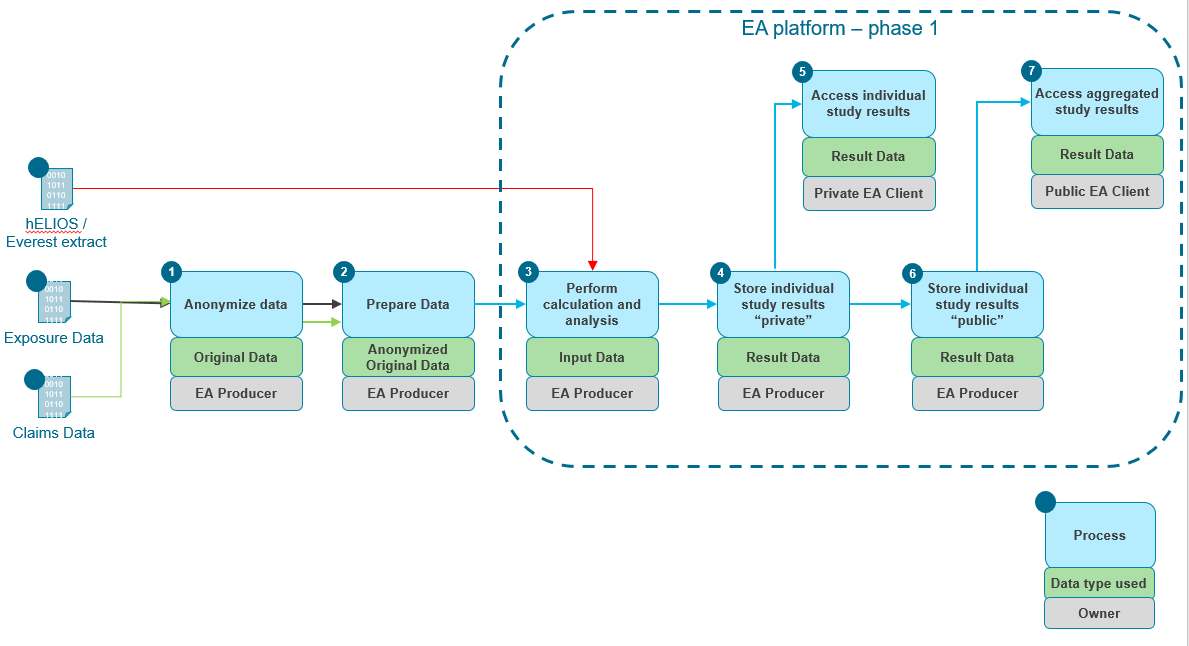
[Annex C – Portfolio-specific assessment 18](#_Toc528256873)

# Introduction to the Experience Analysis Platform

* 1. **Aim**

The aim of this paper is to:

1. explain SGL’s[[1]](#footnote-2) obligations in relation to complying with key legal requirements applicable to the Experience Analysis Platform; and
2. to document the criteria which must be applied when anonymising personally identifiable information prior to uploading the same to the Experience Analysis platform.
   1. **Introduction to the Experience Analysis Platform**
      1. Experience analysis is the process of transforming historical granular portfolio data into standard analysis metrics.
      2. Its main goals are to:
3. support the development of actuarial assumptions, used in various internal processes like pricing, reserving, modelling;
4. facilitate research and development; and
5. provide insights into portfolio characteristics.
   * 1. An experience analysis study corresponds to the analysis of a defined scope (e.g. one reinsurance treaty or the portfolio of one type of risk from a cedant) and involves as main process steps (detailed at paragraph 1.3, below):
6. ***Data preparation***: Cleaning of data received and transformation into a standardised data format.
7. ***Calculation***: Transformation of standardised portfolio data into experience metrics.
8. ***Analysis & reporting***: Analysis of produced results by various dimensions and reporting to internal and/or external clients of the study.
   * 1. Data used for experience analysis typically corresponds to pricing data received during the tender process or reporting data provided by cedants as part of reinsurance accounting.
     2. The Experience Analysis Platform ultimately aims to provide an integrated solution covering the entire above experience analysis process. However, in the first development phase the scope is limited to the calculation and analysis and reporting processes. Data preparation remains outside of the platform.
     3. The main goals of the Experience Analysis Platform are to:
9. Increase automation and user-friendliness of the production of an experience analysis study.
10. Standardise methodologies and result format.
11. Provide easy central access to study results for all interested internal parties, allowing internal Experience Analysis Clients to aggregate results from multiple studies.
12. Reduce operation risks and assure compliance to all contractual and legal obligations.
    1. **The Experience Analysis Process**
       1. The following diagram represents the simplified flow of data through the Experience Analysis Platform focussing on data types potentially containing personal information and their main data processors.



* + 1. **Stage 1**: Seriatim (original) data from ad-hoc studies is received and personally identifiable information is removed in accordance with the rules set forth in paragraph 3.3 and further described at paragraph 3.5, below. Data can be received (i) directly from the external source (cedant) or (ii) provided internally from the original receiver (accountant, pricing actuary) to the Experience Analysis Producer.
    2. **Stage 2**: Data received can be in various formats and across multiple files and needs to be transformed into the standard data structure and format as defined for the Experience Analysis Platform. In addition, data received can contain various data anomalies which need to be corrected by the Experience Analysis Producer. This can entail an iterative process of clarifying data with the client, correcting data and potentially receiving new data versions.

Once the seriatim data has been through this process, it will be considered standardised and anonymised and is referred to as “Input Data” at this point.

* + 1. **Stage 3**: The calculation is performed by the Experience Analysis Producer and results are analysed. This can entail further iterations of data corrections, calculation assumption setting and results analysis.
    2. **Stage 4**: The study results are stored in a database, which is accessible to a defined group, the “Private” Experience Analysis Clients. “Private” hereby relates to the limitation of use of data to the original purpose the data has been received for. Limitations mostly result from reinsurance treaty conditions or other contractual arrangements between SCOR and its client, typically an insurance company.

*h*ELIOS will be built with functionality to mark data for additional use or not. However, a process for non-*h*ELIOS data will be developed, as further described in paragraph 3.5, below. Upon certain events (time event, client request) data might be deleted from the database.

* + 1. **Stage 5**: Experience Analysis Clients access the results via a tool interface where they can analyse the results across any risk dimension available. Access to results is limited to the format of Result Data, which excludes certain detailed information from the Input Data, including any personally identifiable information. Result Data is typically aggregated across policies but may contain individual policy data.
    2. **Stage 6**: In case of no contractual limitations, Result Data is added to a public result repository. Physically this might be the same database and just consist in a flag indicating that result data is accessible as part of the global repository[[2]](#footnote-3).
    3. **Stage 7**: The global repository can be accessed by a wide audience for purpose of statistical analysis. Analysis can be performed on individual or multiple portfolios.
  1. **Data Sources**
     1. The purpose of experience analysis is of statistical nature, requiring a sufficient aggregation of experience to assure credibility of results. ***The results produced by the experience analysis process therefore do not require and do not include personally identifiable information***.
     2. As part of the data preparation process identifiers are used for identifying information which relates to a given insured or policy, for example to be able to link information belonging to the same person or policy across different data files, but these will be anonymised before being uploaded into the Experience Analysis Platform.
     3. Anonymisation is permissible if reidentification is reasonably impossible. Properly anonymising personal data takes it outside of scope for the GDPR.
     4. Experience analysis requires as many risk characteristics as possible, which when used in combination carries a *potential* risk of identifying individuals. To allow a meaningful and credible analysis, any variable will need to be sufficiently aggregated to avoid identification of an individual.
     5. The Experience Analysis Platform will receive data from:

1. in-force portfolio systems (*h*ELIOS / Everest), via a standard interface providing anonymised data;
2. ad-hoc data for example received under pricing studies or provided by cedent for performing analysis as a service; or
3. in very limited circumstances, purchased from external third parties.
   * 1. There is (i) a defined standard data format and (ii) data dictionary, specifying all common variables accepted by the platform with their name, type and in many cases, standardised categories.
   1. **Data Dictionary**
      1. The Experience Analysis Platform standardizes the methodologies for performing studies and calculating results. A fundamental part of this is the standardisation of data used, to allow, among others, an automated process for performing calculations and to allow the accumulation of results across multiple studies based on standardized data definitions.
      2. A standard data dictionary has been defined for both “Input Data” and “Result Data” which includes:
4. Standard names and definitions for all data variables.
5. Type of variables (data / numerical / text), text variables consisting of those with a standard referential of possible values (code variables) and those without referential (free text).
6. A standard data referential for the majority of text variables.

The full data dictionary is given in Annex A.

* + 1. Certain variables are defined as compulsory and must be contained in the Input Data uploaded to the platform, whereas the majority of variables is optional.
    2. The data dictionary is managed centrally by the Global Administration Team of the Experience Analysis Platform. Any data uploaded to the Experience Analysis Platform is checked for complying with the data dictionary and in case of any difference the data file is automatically rejected by the Platform.
    3. The data dictionary has been defined globally based on an exhaustive inventory of all commonly used data variables in all key Experience Analysis markets of SGL. Occasionally clients may provide other types of information relevant to a given study, which is facilitated via dummy variables in the data dictionary. These dummy variables can be filled by the user ad-hoc with his specific information.
    4. If a new type of information is provided more regularly, the information will be added to the standard data dictionary by the Global Administration Team for inclusion to a subsequent platform update. ***In case of addition of a variable relating to personal information, the addition will be subject to a review and approval by the Central Review Team*** (discussed further at [3.2.5], below).
  1. **Ad-hoc Studies**
     1. Apart from data provided by clients as part of their regular reinsurance reporting, which is used by different SCOR departments and managed by dedicated policy systems like *h*ELIOS and Everest, clients may also provide additional data ad-hoc for different other purposes. Most notably data is provided to support the pricing of a new reinsurance treaty, but clients may also provide data for other reinsurance services, like performing benchmark studies across a market.
     2. Whereas data from in-force portfolio systems has already been cleaned, standardised and anonymised for consumption by the Experience Analysis Platform; for ad-hoc studies the data is usually transferred in its original format as provided by the cedant to the Experience Analysis Producer.
     3. Data provided under an ad-hoc pricing study is typically received by local client managers or marketing/pricing actuaries and transferred to Experience Analysis actuaries. Experience analysis is the main usage of the data provided but data may also be used for other pricing processes outside the scope of the Experience Analysis Platform.
     4. Typically, clients will exclude any personally identifiable information from the data provided, but sometimes fail to do so and the removal of such information is a crucial part of the anonymisation process.
  2. **Research and Development**
     1. While most experience analysis studies are primarily performed for a specific purpose, like the pricing of a given portfolio or the reserving of an in-force treaty, study results are also valuable from a research and development perspective and used to improve our risk understanding, like investigating new risk dimensions or to adapt the modelling of actuarial assumptions.
     2. Research and development analysis is based on the same data structure and types of results but may require more granular information on certain variables to allow the investigation of new risk aspects. For example, while a pricing study may group all types of occupations into only ten risk classes, a research and development study might analyse the experience of individual types of occupation to potentially develop a new set of occupation classes.

# Legal Considerations

* 1. **Data Protection Legislation**
     1. The GDPR[[3]](#footnote-4) is directly applicable in all EU Member States without the need for transposition from 25 May 2018. It is a regulation concerning the acquisition, use and disposal of personal data.
     2. Issued by the European Commission and binding on all Member States and all enterprises operating within those Member States. It also applies to organisations outside of the EU handling personal data of European Citizens.
     3. Compliance with the GDPR is mandatory. Failure to comply with the obligations set out within the GDPR could result in financial penalties up to 4% of annual global turnover.
     4. SGL’s baseline position is based on GDPR. It is the foundation on which the Life Division has set its approach to data protection compliance. However, it is recognised that there may be local legal requirements that must be considered.
     5. The Experience Analysis Platform will be built as a central platform on servers within the EU (in France), therefore any studies will be brought within scope for GDPR and will have to comply with the general anonymisation rules set forth in this paper.
     6. The GDPR has an element of extraterritorial reach, meaning that it will apply to controller’s or processor’s based in the EU and also to controllers and processor’s based outside of the EU who: (i) offer goods or services to EU data subjects (individuals) whether or not payment is required, and (ii) monitor the behaviour of such citizens where their behaviour takes place within the EU.
     7. It is worth noting that whilst the GDPR provides a high standard of compliance, many other countries around the world are increasing their local data protection laws and the relevant local legislation should also be considered. This topic will be kept under close review by the SGL Legal team.
     8. Anonymisation is permitted under the GDPR and provided the reinsurance agreement does not explicitly prohibit this activity is a permitted use of the data. It is a type of processing under the GDPR and is specified in the outwards facing fair processing notice. ***Anonymising personal data takes it outside of scope for the GDPR***.
  2. **Processing Personal Data within the Experience Analysis Platform**
     1. The seriatim data that is received by SGL prior to being uploaded into the Experience Analysis Platform may comprise personal data. As mentioned in [1.3] above, such personal data must be anonymised and reformatted into the Experience Analysis standard format.
     2. Personally identifiable information must always be processed in line with applicable data protection laws, including the GDPR.
     3. The act of anonymising personal data is a type of processing under the GDPR. Therefore, one must establish the legal basis on which the data is to be anonymised. In other words, it must either (i) be compatible with the original purpose, for example in direct connection with the performance of an insurance contract or (ii) that the anonymisation is in SCOR’s interest. With the latter, SCOR is under an obligation to inform data subjects and fulfils this obligation with an appropriate statement on the external website, www.scor.com.
     4. However, as mentioned above, the data used in the Experience Analysis Platform will be anonymised and once anonymised will fall outside the scope of the GDPR.

# Anonymisation Guidelines

* 1. **What is anonymisation**
     1. To anonymise personal data for use in the Experience Analysis Platform is to remove identifying particulars or details from the seriatim data for statistical or other purposes including for use on the Experience Analysis Platform rendering the reidentification of individuals impossible. In other words, it is the irreversible transformation of data that prevents re-identification. Therefore, personal data is anonymous if re-identification is reasonably impossible.
     2. This should be distinguished from the concept of pseudonymisation which is a recognised security measure and which can be described as the reversable transformation of data i.e. the obfuscated identifier can be mapped back to the original identifier (but must be kept separately).
     3. Therefore, one must make the distinction between anonymisation which takes personal data outside the scope of the GDPR and Pseudonymisation which retains the personally identifiable characteristics thus remains within scope for GDPR.
     4. ***The Experience Analysis Platform will use anonymised personal data***. Data will be received either from (i) the *h*ELIOS Platform or (ii) other sources. In-force data received from *h*ELIOS will be anonymised when output from that tool, prior to being uploaded into Experience Analysis. Whereas the Experience Analysis User will be required to manually anonymise data prior to uploading to Experience Analysis. In both cases according to the anonymisation rules set forth in this paper.
     5. The act of anonymising personal data is a form of processing and up to this point will need to comply with relevant data protection laws including the GDPR. Provided (i) there are no conflicting contractual restrictions and (ii) SCOR has complied with its legal obligations[[4]](#footnote-5), it can anonymise the personal data either:

1. on the basis that it is necessary for the performance of the underlying reinsurance agreement where the anonymisation is required to perform experience analysis studies directly connected to an identified reinsurance relationship; or
2. pursuant to SCOR’s legitimate interest in performing studies to improve pricing assessments or for research and development.
   1. **Anonymisation assessment**
      1. The anonymization assessment is performed independently of the source of the data.
      2. The identification of individual persons can be achieved via unique-identifiers or the combination of multiple quasi-identifiers. Quasi-identifiers are pieces of information that are not themselves unique-identifiers, but which are sufficiently well correlated with a person that they can be combined with other quasi-identifiers to create a unique identifier (source: Wikipedia).
      3. In all circumstances the data may not contain any unique-identifiers of personal information, in particular any of the following information:

* Forename Name
* Middle Name(s)
* Surname
* Abbreviation of names and/or initials
* Social Security Number / National Insurance Number or similar
* Passport or National ID number
* Policy ID, unless anonymized by the in-force portfolio system or by the user for other data sources.
  + 1. Two options for the assessment of quasi-identifiers are provided, (i) a standard set of rules and (ii) a portfolio specific assessment. At least one of the two must be performed by the user and compliance has to be confirmed, assuring that data is anonymous before being uploaded onto the platform.
    2. Any data not being sufficiently anonymised based on the above assessment criteria may be referred to the Central Support Team for an individual review.
    3. The Central Support Team will consist of the Experience Analysis Project Manager, SGL Legal [and the SCOR Group Data Protection Officer] who are tasked with balancing legal obligations and the business request. They will be tasked with assessing what, if any additional options may be considered to reduce the risk of identification. It is also recognized that inherent characteristics of the data may reduce the real risk of identification, like its quality or its age. It is recognised that the quality of the data may change over time. The anonymisation guidelines are not a one-off and will be reviewed by the Central Support Team on a three yearly basis taking into account advances in technology and anonymisation techniques.
    4. A standard set of data rules has been defined which can be applied by users for standard types of business and types of studies to assure the data is compliant. These rules apply to all relevant personal data information and define a stricter maximum level of granularity for each variable to assure anonymisation even in presence of all variables.
    5. For certain types of studies or non-standard types of datasets, users will perform their own assessment based on defined rules to tailor the anonymisation to the actual information contained in the given dataset.
  1. **Standard Rules**
     1. All data uploaded must follow the standard data dictionary defined. No personal related information may be entered in the dummy variables (or any other variables). This assures that no identifiers or other quasi-identifiers are contained in the data.
     2. The dummy variables may be used for other types of information, as long as they are clearly not related to personal information, like certain information on the insurance product characteristics or on the cedent.
     3. The data dictionary has been assessed for any quasi-identifiers i.e. any fields that when used in combination, may lead to identifying an individual. Thus, the data dictionary was assessed balancing the opposing needs of the business to retain sufficiently detailed data sets versus the legal need to irreversibly prevent identification of individual data subjects. It is recognised that certain fields are more obvious to obfuscate than others. The following quasi-identifiers have been identified, for which additional restrictions apply:

|  |  |  |
| --- | --- | --- |
| **Category of information** | **EA variable** | **Rule** |
| Demographic | Date of Birth | Restricted to monthly format[[5]](#footnote-6) |
| Geographic | Region of Residence | Following standard data dictionary referential if existent for given country, otherwise at most 20 categories |
| Occupational | Salary Band | At most 10 categories |
| Educational | Education level | Not accepted |

Occupation class is not considered a quasi-identifier as it is a numerical representation of the occupation classification performed by the insurer. It is specific to the insurer and not public and of no value for identification purpose as one class typically represents a wide mix of very diverse types of occupations.

Certain medical information such as BMI and cause of claim are not considered quasi-identifiers. All categories must be sufficiently homogenous.

The justification of the above rules is provided in Annex B

Data which does not meet this criterion will be not be considered as anonymized unless concluded so under the Portfolio specific assessment defined under 3.4.

* 1. **Portfolio specific assessment**

The portfolio specific assessment takes account of the variables actually contained in the dataset with their corresponding granularity.

In line with the general rule, no unique identifiers may be contained in the dataset.

For the assessment of quasi-identifiers all variables related to the following categories of information have to be included in the assessment:

|  |  |
| --- | --- |
| **Category of information** | **Variable examples** |
| Demographic | Gender  Date of Birth  Age at commencement |
| Geographical information | Region of Residence  Postal code |
| Occupational information | Occupation title  Occupation groups  Salary  Salary Band |
| Educational information | Education level  Years of education completed |

Based on the demographics of the country in scope, the minimum number of persons per cell is calculated based on the most granular variable for each category of information.

The final number of Persons per cell has to be [2] or more.

The number of persons of value v should, whenever possible, be based on the actual distribution of the country in scope. In absence of reliable population statistics, the actuary will prepare an estimation and document its justification.

A template for conducing the assessment is provided in Annex C. Examples are also provided in Annex B.

# Annex A – Data dictionary



# Annex B – Standard Rule

To validate the proposed set of rules, it is tested on different key markets of SGL.

The rules are applied to the general population of the given market, estimating the number of persons contained in each cell, defined by the combination of the information of all quasi-identifiers.

The assessment follows the method outlined in the “portfolio specific assessment”. For each category of information, the reduction of the number of persons in scope due to the additional information is estimated.

For example, for Region of Residence the number of persons of the smallest Region is divided by the total population, representing the maximum reduction in cell size due to this information.

Wherever possible the distribution of variables is based on actual census data.

## Test 1: UK

The age exposure is set to 18 – 70, representing the large majority of insurance exposure.

|  |  |  |
| --- | --- | --- |
| Information category | Variable | Source |
| Demographic | Gender & Age | 2017 Population estimate by UK Office for National Statistics, split by gender, individual age and region |
| Geographic | Region of Residence |
| Occupational | Salary Band | Assuming a reasonably homogenous distribution in line with industry practice. The smallest category assumed to represent 5% of total persons in portfolio. |

|  |  |  |
| --- | --- | --- |
|  | Minimum cell size | Total cell size |
| Total population |  | 54 786 327 |
| Gender & age & region | 13 272 | \* 0.024% |
| Month of Birth | 1/12 of population | \* 8.3% |
| Occupational | 5% of population | \* 5% |
| Cell size |  | 55 |

After taking into account all information from gender, age, month of birth, region and occupational information, the smallest cell of characteristics still contains 55 persons.

The calculation details can be found in attached file: 

## Test 2: France

The age exposure is set to 18 – 70, representing the large majority of insurance exposure.

|  |  |  |
| --- | --- | --- |
| Information category | Variable | Source |
| Demographic | Gender & Age | 2017 population statistics by INSEE (national institute of statistics) split by gender & individual age |
| Geographic | Region of Residence | 2017 population statistics by INSEE (national institute of statistics), by region |
| Occupational | Salary Band | Assuming a reasonably homogenous distribution in line with industry practice. The smallest category assumed to represent 5% of total persons in portfolio. |

|  |  |  |
| --- | --- | --- |
|  | Minimum cell size | Total cell size |
| Total population |  | 67 186 638 |
| Gender & age | 348 020 | \* 0.52% |
| Month of Birth | 1/12 of population | \* 8.3% |
| Region | 337 796 | \* 0.50% |
| Occupational | 5% of population | \* 5% |
| Cell size |  | 7 |

After taking into account all information from gender, age, month of birth, region and occupational information, the smallest cell of characteristics still contains 7 persons.

The result is strongly impacted by the small region of Corsica, which only contains around 340’000 people, compared to the next larger region with more than 2m people. Without Corsica, the minimum total cell size would be 47.

The calculation details can be found in attached file:



## Test 3: Germany

The age exposure is set to 18 – 70, representing the large majority of insurance exposure.

|  |  |  |
| --- | --- | --- |
| Information category | Variable | Source |
| Demographic | Gender & Age | 2011 population census, split by gender & individual age |
| Geographic | Region of Residence | 2011 population census, split by region |
| Occupational | Salary Band | Assuming a reasonably homogenous distribution in line with industry practice. The smallest category assumed to represent 5% of total persons in portfolio. |

|  |  |  |
| --- | --- | --- |
|  | Minimum cell size | Total cell size |
| Total population |  | 80 209 997 |
| Gender & age | 292 525 | \* 0.36% |
| Month of Birth | 1/12 of population | \* 8.3% |
| Region | 650 863 | \* 0.81% |
| Occupational | 5% of population | \* 5% |
| Cell size |  | 10 |

After taking into account all information from gender, age, month of birth, region and occupational information, the smallest cell of characteristics still contains 10 persons.

The calculation details can be found in attached file:



# Annex C – Portfolio-specific assessment

The assessment can be performed using attached template.



1. In this paper “SGL” refers to SCOR Global Life SE and its subsidiaries. [↑](#footnote-ref-2)
2. The structure of this result database is under consideration by the project team. [↑](#footnote-ref-3)
3. Regulation (EU) 2016/679 of the European Parliament and of the Council, the “General Data Protection Regulation” or the “GDPR” [↑](#footnote-ref-4)
4. Which include but are not limited to referring to this in the outwards facing fair processing notice which can be found at [www.scor.com](http://www.scor.com) and applying appropriate technical and organisational measures. [↑](#footnote-ref-5)
5. Technically may be entered as »15/MM/YYYY » [↑](#footnote-ref-6)