Draft Project plan for the CEN Workshop on "the QaR method to measure the extreme risk of re-identification of a database in the context of assessing its insurability"

Requests to participate in the Workshop and/or comments on the project plan are to be submitted by 20/04/2023 to aylin.kip@afnor.org[1]

Recipients of this project plan are kindly requested to name all patent rights known to them to be relevant to the Workshop and to make available all supporting documents.

Paris, 07/03/2023 (Version 1.0)

---

[1] Applications for participating in the Workshop and comments on the project plan that are not received by the deadline do not need to be taken into consideration. Once constituted, the Workshop will decide whether or not to consider the comments received in good time.

Contents

Summary

Various anonymisation and pseudonymisation techniques exist, some of which are standardized. Nevertheless, the risk of a data breach for anonymised data and therefore that of re-identification of individuals remain significant. Therefore, institutions and commercial companies that may create or disseminate databases without identifiers are thus looking for a simple tool to measure this extreme risk of re-identification.

The shortcomings of current techniques associated with the regulations on personal data protection bring out a particular need: to propose a more practical method to measure the efficiency of the techniques implemented to protect anonymised personal data from the risk of re-identification.

The Quantile at Risk (QaR) method provides an innovative solution to this quandary by providing a continuous assessment of the risk under consideration. It fills the gap between anonymised data and personal data, - including pseudonymised data - created by regulation on personal data conditioned primarily by compliancy.

The planned Workshop specifies a method to measure the quality of the anonymization of a data file, called QaR (Quantile at Risk) and how the QaR method results should be fine-tuned to data from different sectors (health care, banking, telecommunication, etc.). It is applicable to any database containing personal data and is intended to be used by all stakeholders processing personal data in any sector (e.g. healthcare, banking, telecommunication). It does not apply to databases containing among their inputs an individual identifier or transformed by hashing or other mathematical data transformation technics.

1 Status of the project plan

Draft project plan for public commenting (Version 1.0)

This draft project plan is intended to inform the public of a new Workshop. Any interested party can take part in this Workshop and/or comment on this draft project plan. Please send any requests to participate or comments by e-mail to **aylin.kip@afnor.org**.

All those who have applied for participation or have commented on the project plan by the deadline will be invited to the kick-off meeting of the Workshop on **20-04-2023**.

2 Workshop proposer and Workshop participants

2.1 Workshop proposer

Person or organisation	**Short description and interest in the subject**
Quantos SA	Quantos aims to position data analytics in business reality and continuously provide innovative solutions and consulting.

2.2 Other potential participants

This CWA will be developed in a Workshop (temporary body) that is open to any interested party. The participation of other experts would be helpful and is desired. It is recommended that:

- organizations specializing in statistics
- organizations specializing in data protection laws

- organizations specializing in data analytics
- research institutions working on data and privacy
- any organization that uses pseudonymized or anonymized data

take part in the development of this CWA.

3 Workshop objectives and scope

3.1 Background

Big data consists of collecting and analysing datasets with ever more sophisticated algorithms. These datasets may contain personal data, the processing of which is strictly regulated worldwide, especially in the European Union since the entry into force of the General Data Protection Regulation (GDPR) on May 25, 2018.

In order to comply with obligations required by such regulations, namely with Article 32 "Security of processing" and Article 35 "Data protection impact assessment" of the GDPR, data controllers may rely on anonymisation or pseudonymisation of large-scale data to remove any information that may relate directly or indirectly to an individual.

Anonymisation results from processing personal data in order to irreversibly prevent identification whereas pseudonymisation is a reversible method that reduces the possibility of linking a dataset with the original identity of a data subject, which makes it a useful security measure. The second option offers the possibility to reidentify a data subject thanks to additional information stored separately by data controllers or their data processors.

Consequently, the data controller who directly or indirectly anonymises or pseudonymises personal data must be able to demonstrate, in the event of a complaint, audit or data breach, that, in accordance with articles 32 and 35 of the GDPR, it has implemented the appropriate technical and organisational means, taking into account the state of the art and the degree of sensitivity of the data, and the associated risks, to ensure that the pseudonymisation or anonymisation process that has been carried out allows for preventing the reidentification of data subjects.

Once personal data have been anonymised, the GDPR no longer applies to the dataset, since, in theory, it no longer contains personal data, irreversibly so. However, this is not the case for the processing of pseudonymised data, which is by nature reversible, and to which the GDPR remains applicable.

Nonetheless, given that the state of the art evolves continually, and with it, the risk of re-identification of individuals, pseudonymisation or anonymisation of personal data is necessarily an iterative process, involving constant and periodic monitoring of the new possibilities of re-identification offered by technological progress and the new data available allowing new data cross-references[2].

Amongst technological advances, a data controller proceeding directly or indirectly to the pseudonymisation or anonymisation of personal data must in particular take into account those that increase the risk of re-identification, but also the new tools that make it possible to verify the relevance of the initial pseudonymisation or anonymisation risk assessment.

Various anonymisation and pseudonymisation techniques exist, some of which are standardized. Nevertheless, the risk of a data breach for anonymised data and therefore that of re-identification of individuals remain significant.

Institutions and commercial companies that may create or disseminate databases without identifiers are thus looking for a simple tool to measure this extreme risk of re-identification.

Preventive data anonymisation methods require to employ substantial and expensive IT resources due to the algorithmic complexity of calculations.

---

[2] Opinion 05/2014 on Anonymisation Techniques: "*Data controllers should consider that an anonymised dataset can still present residual risks to data subjects. Indeed, on the one hand, anonymisation and re-identification are active fields of research and new discoveries are regularly published, and on the other hand even anonymised data, like statistics, may be used to enrich existing profiles of individuals, thus creating new data protection issues. Thus, anonymisation should not be regarded as a one-off exercise and the attending risks should be reassessed regularly by data controllers.*"

The shortcomings of current techniques associated with the regulations on personal data protection bring out a particular need: to propose a more practical method to measure the efficiency of the techniques implemented to protect anonymised personal data from the risk of re-identification.

The Quantile at Risk (QaR) method provides an innovative solution to this quandary by providing a continuous assessment of the risk under consideration. It fills the gap between anonymised data and personal data, - including pseudonymised data - created by regulation on personal data conditioned primarily by compliancy.

The QaR method that measures the extreme risk of re-identification of a dataset responds to challenges arising from evolutions of Big Data and artificial intelligence. Any sector confronted with the risk of re-identification of the personal data they process, including but not limited to the health sector, could benefit from this innovation.

To promote and experiment this method on a larger scale, it is proposed to develop a CWA (CEN Workshop Agreement) using as basis the AFNOR Spec Z90-030 "The QaR method for measuring the extreme risk of re-identification of a database in the context of assessing its insurability" published in February 2020. A CEN Workshop will allow stakeholders from various sectors to share their needs and feedbacks in a transparent and open environment. This deliverable will give the opportunity to make available to a larger audience an innovative method to assess the risk of re-identification of a database.

According to the feedbacks about the CWA, a European or international standardization process could be undertaken.

3.2 Scope

The planned Workshop specifies a method to measure the quality of the anonymization of a data file, called QaR (Quantile at Risk) and how the QaR method results should be fine-tuned to data from different sectors (health care, banking, telecommunication, etc.).

The planned Workshop is applicable to any database containing personal data.

The planned Workshop is intended to be used by all stakeholders processing personal data in any sector (e.g. healthcare, banking, telecommunication).

The planned Workshop does not apply to databases containing among their inputs an individual identifier or transformed by hashing or other mathematical data transformation technics.

3.3 Related activities

The subject of the planned CWA is not at present the subject of a standard. However, there are committees, standards and/or other technical specifications that deal with related subjects and thus need to be taken into account - and involved, where necessary - during this Workshop:

- ISO/IEC JTC 1/SC 27
  - ISO/IEC 20889:2018 Privacy enhancing data de-identification terminology and classification of techniques
  - ISO/IEC 27559:2022 Information security, cybersecurity and privacy protection – Privacy enhancing data de-identification framework
- CEN/CLC JTC 13

See Annex A "Bibliography" for related scientific publications.

4 Workshop programme

4.1 General

The kick-off meeting is planned to take place on 20-04-2023 in Paris, France and virtually via Zoom. A draft for public commenting is planned to be published for 60 days on November 2023.

A total of four Workshop meetings (kick-off meeting and Workshop meetings) will be held, during which the content of the CWA will be presented, discussed and approved. Meetings will be held remotely unless agreed otherwise.

The CWA will be drawn up in English (language of meetings, minutes, etc.). The CWA will also be written in English.

4.2 Workshop schedule

Table 1: Workshop schedule (preliminary)

CEN/CENELEC Workshop	M01	M02	M03	M04	M05	M06	M07	M08	M09	M10	M11	M12	...
Initiation	███	███	███										
1. Proposal form submission and TC response	▓	▓											
2. Project plan development		▓											
3. Open commenting period on draft project plan (mandatory)		▓	▓										
Operation			███	███	███	███	███	███	███	███	███		
4. Kick-off meeting			▓										
5. CWA(s) development			▓	▓	▓	▓	▓	▓					
6. Open commenting period on draft CWA(s) (optional)									▓	▓			
7. CWA(s) finalised and approved by Workshop participants											▓		
Publication												███	
8. CWA(s) publication												▓	
Dissemination (see 7)		▓	▓					▓	▓	▓		▓	▓
Milestones			K	V	V	M	V				M/A	P	D

B   CEN/CENELEC BT meeting deciding on establishment of a CEN/CENELEC Workshop
K   Kick-off
M   Workshop meeting
V   Virtual Workshop meeting
A   Adoption of CWA
P   Publication of CWA
D   Online distribution of CWA

5   Resource planning

The Workshop will be financed by the participants and the Workshop proposer. The participation fee is 3.000 euros. Universities can participate without making a financial contribution.

6   Workshop structure and rules of cooperation

6.1 Participation in the Workshop

The Workshop will be constituted during the course of the kick-off meeting. By approving this project plan, the interested parties declare their willingness to participate in the Workshop and will be formally named as Workshop participants, with the associated rights and duties. Participants at the kick-off meeting who do not approve the project plan are not given the status of a Workshop participant and are thus excluded from further decisions made during the kick-off meeting and from any other decisions regarding the Workshop.

As a rule, the request to participate in the Workshop is closed once it is constituted. The current Workshop participants shall decide whether any additional members will be accepted or not.

Any new participant in the Workshop at a later date is decided on by the participants making up the Workshop at that time. It is particularly important to consider these aspects:

a. expansion would be conducive to shortening the duration of the Workshop or to avoiding or averting an impending delay in the planned duration of the Workshop;
b. the expansion would not result in the Workshop taking longer to complete;
c. the new Workshop participant would not address any new or complementary issues beyond the scope defined and approved in the project plan;
d. the new Workshop participant would bring complementary expertise into the Workshop in order to incorporate the latest scientific findings and state-of-the-art knowledge;
e. the new Workshop participant would actively participate in the drafting of the manuscript by submitting concrete, not abstract, proposals and contributions;
f. the new Workshop participant would ensure wider application of the CWA.

All Workshop participants who voted for the publication of the CWA or its draft will be named as authors in the European Foreword, including the organisations which they represent. All Workshop participants who voted against the publication of the CWA, or who have abstained, will not be named in the European Foreword.

6.2 Workshop responsibilities

The Workshop Chair is responsible for content management and any decision-making and voting procedures. The Workshop Chair is supported by the Workshop Vice-Chair and the responsible Workshop secretariat, whereby the Workshop secretariat will always remain neutral regarding the content of the CWA(s). Furthermore, the Workshop secretariat shall ensure that CEN-CENELEC's rules of procedure, rules of presentation, and the principles governing the publication of CWA(s) have been observed. Should a Workshop Chair no longer be able to carry out her/his duties, the Workshop secretariat shall initiate the election of a new Workshop Chair. The list below covers the main tasks of the Workshop Chair. It is not intended to be exhaustive.

– Content related contact point for the Workshop
– Presides at Workshop meetings
– Ensures that the development of the CWA respects the principles and content of the adopted project plan
– Manages the consensus building process, decides when the Workshop participants have reached agreement on the final CWA, on the basis of the comments received
– Ensures due information exchange with the Workshop secretariat
– Represents the Workshop and its results to exterior

The Workshop secretariat, provided by a CEN/CENELEC national member, is responsible for organising and leading the kick-off meeting, in consultation with the Workshop proposer. Further Workshop meetings and/or web conferences shall be organised by the Workshop secretariat in consultation with the Workshop Chair. The list below covers the main tasks of the Workshop secretariat. It is not intended to be exhaustive.

– Administrative and organisational contact point for the Workshop

- Ensures that the development of the CWA respects the principles and content of the adopted project plan and of the requirements of the CEN-CENELEC Guide 29
- Formally registers Workshop participants and maintains record of participating organisations and individuals
- Offers infrastructure and manage documents and their distribution through an electronic platform
- Prepares agenda and distribute information on meetings and meeting minutes as well as follow-up actions of the Workshop
- Initiates and manage CWA approval process upon decision by the Workshop Chair
- Interface with CEN-CENELEC Management Centre (CCMC) and Workshop Chair regarding strategic directions, problems arising, and external relationships
- Advises on CEN-CENELEC rules and bring any major problems encountered (if any) in the development of the CWA to the attention of CEN-CENELEC Management Centre (CCMC)
- Administrates the connection with relevant CEN or CENELEC/TCs

6.3 Decision making process

The CEN/CENELEC Workshop Chair shall decide when the CEN/CENELEC Workshop participants have reached agreement on the final text of the CWA, on the basis of the comments received and any further consultation that has taken place, at which point the CWA is approved.

7   Dissemination and participation strategy
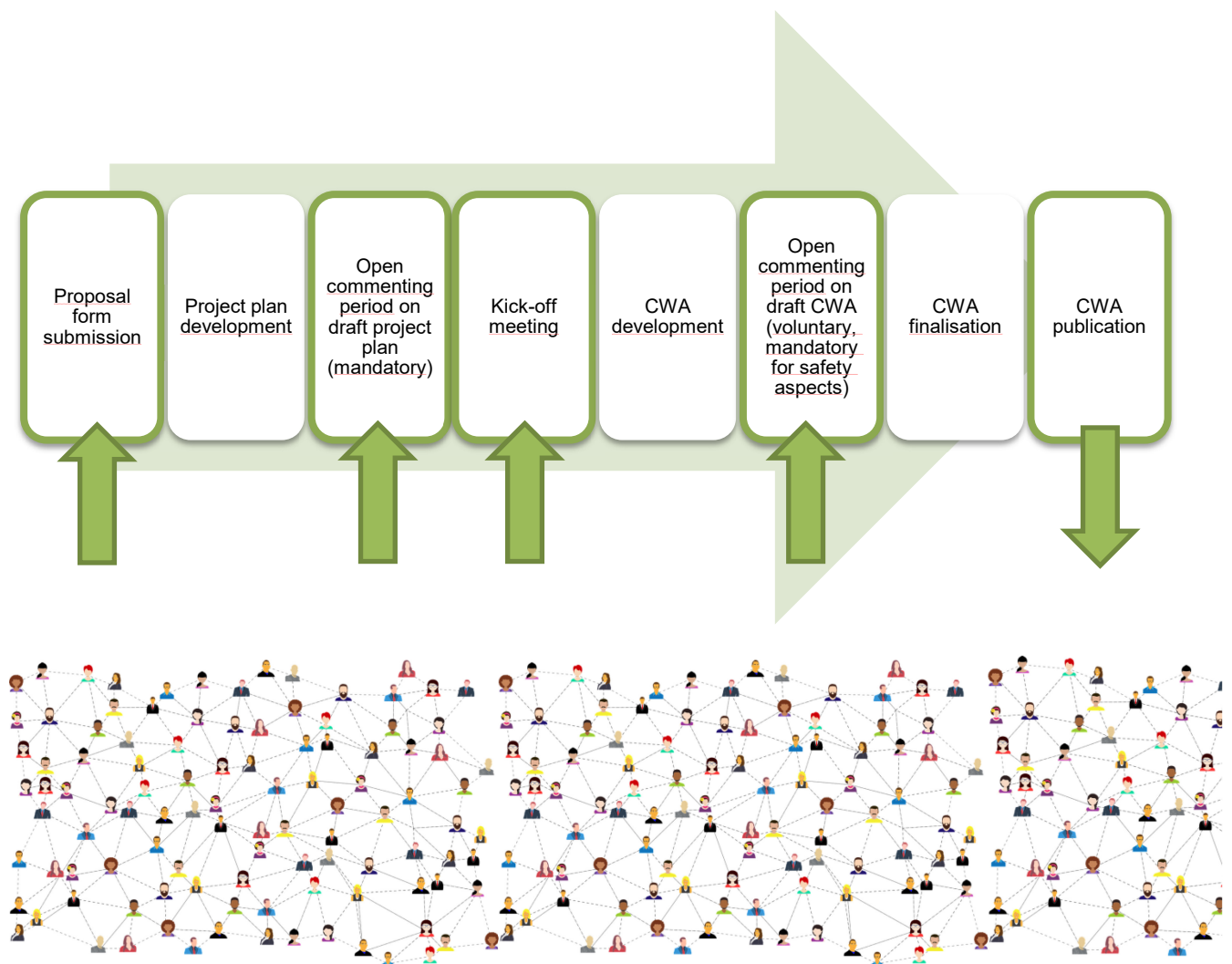
CWA publication

In addition to the CCMC website, the published CWA will be advertised on social media.

8   Contacts

– Workshop Chair:

Michel Bera
CNAM (Conservatoire national des arts et métiers)
292 rue Saint-Martin
F-75141 Paris Cedex 03
+ 33 (0)1 40 27 20 00
michel.bera@lecnam.net
https://www.cnam.fr/

– Workshop Vice-Chair:

Photis Stavropoulos
Quantos SA
154 Syngrou Ave, Kallithea, 17671, Greece
+30 210 9238185-7
photis.stavropoulos@quantos-stat.com
https://quantos-stat.com/

– Workshop Secretariat:

Aylin Kip
AFNOR
11 Rue Francis de Pressensé, 93210 Saint-Denis, France
+33 (0)1.41.62.63.96
aylin.kip@afnor.org
www.afnor.org

– CEN-CENELEC Management Centre:

Laurens Hernalsteen
CCMC
Rue de la Science 23
B - 1040 Brussels, Belgium
+32 25500810
lhernalsteen@cencenelec.eu
https://www.cencenelec.eu/Pages/default.aspx

– Workshop proposer:

Dr. Antonis SPINAKIS
QUANTOS SA
154 Syngrou Ave, Kallithea, 17671, Greece
+30 210 9238185-7
anspi@quantos-stat.com
https://quantos-stat.com/

ANNEX A – BIBLIOGRAPHY

❖ **Conferences**:
- Bera M., Spiliopoulos K., Spinakis A., Stavropoulos P. Eds, "Measuring reidentification risk in (pseudo)anonymized datasets: the QaR method and software", Privacy in Statistical Databases, September 2022, Paris.
- Bera M. Ed., "Quantile at Risk (QaR): an open norm (AFNOR-SPEC) for measuring the quality of anonymisation (reidentification risk) from an actuarial perspective for Insurance", DWIH Conference Tokyo, German Centre for Research and Innovation, December 2019, Tokyo.

❖ **Publications**:
- Bera, M. (June 15, 2020), "Peut-on mesurer la qualité d'anonymisation d'un jeu de données anonymisé ?"[3], *L'Actuariel*, no. 37, https://www.lactuariel.fr/metier/recherche/peut-on-mesurer-la-qualite-danonymisation-dun-jeu-de-donnees-anonymise/
- Bera, M., "Big Data Rapid Re-identification Risk Assessment Method", *in* Planchet, F., Robert, C. Y. Eds, *Insurance Data Analytics*, Paris, Economica, "Assurance Audit Actuariat", 2020, Chapter 14.

❖ **Standardization**:
- Standards developed by ISO/IEC JTC 1/SC 27 "Information security, cybersecurity and privacy protection"
- ISO/IEC 20889:2018 Privacy enhancing data de-identification terminology and classification of techniques
- ISO/IEC 27559:2022 Information security, cybersecurity and privacy protection – Privacy enhancing data de-identification framework

---

[3] Can we measure the quality of anonymisation of an anonymised dataset?