ORIGINAL ARTICLE

From data to decisions – Quality assurance in radon policy

Peter Bossew* and Eric Petermann

German Federal Office for Radiation Protection, Berlin, Germany

Abstract

Radon abatement policy is the response to the detrimental effect of indoor radon which is estimated to cause hundred thousands of lung cancer fatalities worldwide annually. The policy consists of decisions to implement measures. Decisions rest on data and (sometimes competing) interests, among them health protection. Its weight as an argument depends, among other factors, on knowledge about its subject – in this case, levels, effects, and geographical distribution of exposure to radon. Therefore, the quality assurance of radon policy depends on one of the underlying knowledge, from data to decisions derived from them. Some aspects of the quality assurance chain are discussed in this article.

Keywords: *radon policy; quality assurance; decision making; stakeholder interests*

Exposure to indoor radon (Rn) is rated among the foremost causes of lung cancer (1). Being a health risk, regulation is imposed aiming to reduce it. Regulation implies action, laid down in Rn Action Plans, on which the Rn reduction policy is based. Policy means deciding about actions to be taken, and it is controlled by interests that can be competing or even antagonistic; in the case of Rn policy, the interest in protecting the health of the population can be in conflict with an economic interest, because Rn abatement costs money. The pathway from the 'sphere of interests' which leads to legislation (regulation) that contains Rn Action Plans and its implementation in the form of decisions is shown in Fig. 1 as a rough flow scheme.

Some examples of competing interests:

1. In an area, the probability that indoor Rn concentration exceeds a reference level (RL) is above a threshold deemed sufficient to initiate certain action. On the contrary, economical constraints are considered such that the action cannot be entirely fulfilled.
2. Often in historical buildings (e.g. castles and churches), which are workplaces, high Rn concentration occurs due to ancient building style. Remediation is almost impossible because preservation requirements and structural stability do not allow constructional modification. (Reduction of occupation time is not always possible.)
3. According to indoor Rn concentrations, a district should be declared an area, in which Rn prevention and mitigation measures should be legally enforced. The administration fears public opposition and is afraid of losing the next elections.

Focusing on Rn, the objective of radioprotection is to reduce the harm or detriment inflicted by exposure to Rn, that is, to reduce the number of Rn-caused lung cancer cases. In its struggle with competing interests, the stronger the position of radioprotection is, the better its arguments are. This means that they are supported by correct data, scientific knowledge, and competence with proven reliability. The same applies to the proposals on decisions about measures aimed at reducing the detriment caused by exposure to Rn. A necessary (although not sufficient) condition of an Rn action plan to be legally defendable and acceptable to the public is that it is scientifically proven.

The correctness of the links of the chain from data to decisions shall be guaranteed by quality assurance (QA) procedures. Therefore, we speak about the *QA chain*. We propose to distinguish four basic links, namely, *Design QA, Data QA, Evaluation QA,* and *Decision QA*, to be discussed in more detail in this article. In particular, we shall expound on the QA of decisions since it has been granted less attention in the past.

The relevance of QA in Rn policy may be exemplified by these issues:

- Rn policy can be expensive:
  - Surveys are expensive (this does not apply to measurement in the first place, as today detectors and evaluation are cheap; on the other hand, logistics can be expensive: generating a representative sampling design, distribution and collection of large numbers of detectors).

Citation: Journal of the European Radon Association 2022, **3**: 7886 http://dx.doi.org/10.35815/radon.v3.8212

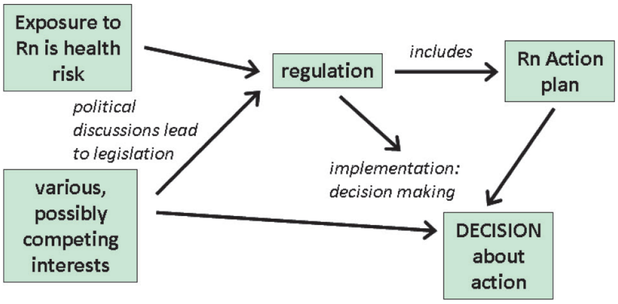

Fig. 1. Pathway from interests to decisions.

  - Remediation of a building is expensive for the owner or for an administration that subsidizes it.
  - Ill-assignment of the status as Rn priority area (RPA[1]) to an administrative unit can be expensive, because the action that is required in an RPA can be expensive.
- Ill-assignment of RPA status can have adverse effects on property value and on psychological factors that may entail political costs.
- Wrong decisions can lead to lawsuits that, very obviously, responsible administrations want to avoid. For example, in several cases, the assignment of RPAs has been challenged in Germany.

The motivation of the thoughts presented here is a debate during the implementation of the European Basic Safety Standards [BSS, (2)] (obligatory for the EU Member States to transpose into National Law) and the European research project MetroRADON (3) (2016–2020, devoted to QA aspects of Rn policy).

Decisions

Formally, decision means selecting one of several possible options for action, based on decision predictors or data (in a wider sense, see below) by applying decision rules or criteria. Hence, a decision can be understood as a triplet (predictors, rules, and possible action), where decision rules map the predictor space into the action space. Predictors or data can be measurement results (e.g. the outcome of an Rn survey), environmental or sociological data that are believed to be relevant for decision-making (e.g. geological maps, geochemical data, population density, climatic data, urbanization, and many other), and 'soft data' such as economic and political interests and constraints (budgetary or with regard to institutional capacity) or public attitude. While the integration of such soft data is difficult, in real cases they can play rather dominant roles.

An intuitive model of decision-making has been proposed by Guo (4). The following scheme is taken from its Wikipedia entry (4) (comments in squared brackets by the authors):

1. Define the problem
2. Establish or Enumerate all the criteria (constraints) [predictors as defined above]
3. Consider or Collect all the alternatives [set of options for action]
4. Identify the best alternative [apply decision rules]
5. Develop and implement a plan of action
6. Evaluate and monitor the solution and examine feedback when necessary.

Steps 3 to 5 correspond to the decision process as introduced earlier. Step 6 is especially important as posterior QA: evaluating whether or to which extent the decision has contributed to solving the problem identified in step 1 (e.g. that detriment inflicted to society by Rn exposure should be reduced), whether there are unwanted side effects, whether resources were used economically, etc. Step 5 should include anterior QA, namely, assessment of the reliability of a decision (to be discussed further below) and probable consequences of a wrong decision.

The quality assurance chain

We propose to divide the QA chain into four links; however, they may intersect, and perhaps more appropriately, one may even speak of a QA network or 'QA rhizome' if one prefers the nonhierarchical approach of Deleuze and Guattari (5). The idea of the QA chain in Rn science has first (to our knowledge) been proposed in the EURAMET project MetroRADON (3).

Design QA

Design QA is about the capability of an effort, such as an experiment, a measurement, or a survey, to enable a targeted decision.

1. If the effort consists in an Rn survey, it must be set up such that a geographical or temporal tendency can be identified reliably (scores to be set), if it exists, while avoiding spurious effects or tendencies.
2. If the task consists in detecting differences or changes (e.g. of mean Rn levels before and after remediation measures), the experiment must be set up such that an effect of a given size can be detected with set confidence, if it exists. Spurious effects shall be avoided with set confidence.

---

[1] The RPA concept underlies the European BSS (2), specifically Art. 103/3, which states that as part of Rn action plans, '*Member States shall identify areas where the radon concentration (as an annual average) in a significant number of buildings is expected to exceed the relevant national reference level*' and Annex XVIII (6), which among items of the Rn action plans names '*Strategy for reducing radon exposure in dwellings and for giving priority to addressing the situations identified under point 2*' (referring to RPAs).

These requirements are related to the *power* of a test, which denotes the probability that an effect is detected, if it exists. Power equals (1-prob) that an effect is missed, although it exists, or 1 – second kind error probability. In a rigid test or experiment, the desired power should be set beforehand, and the test or experiment is designed such that this is fulfilled. While this is difficult in real-world Rn policy, it should at least approximately be attempted, usually through pilot projects.

Specifically in Rn surveying, this entails the following issues:

- *Sample size*: how many locations and how many detectors are necessary?
- Criteria for where and when to locate detectors.
- Assuring representativeness of the sample. (A sample is representative, if its statistical distribution equals the true distribution of the sampled quantity.)
- *Adequate exposure time*: low exposure time leads to high temporal uncertainty with regard to a long-term mean which is usually the target quantity, while experience shows that long exposure leads to increased rate of detector loss.
- Which metadata are needed for the objective of a survey, such as building properties, geo-referencing, and data about inhabitants?
- *How to deal with constraints*: limited budget, limited laboratory resources, or data protection which increasingly turns out an obstacle in Rn policy.

These are severe and nontrivial challenges. Careful planning is therefore advised to avoid suboptimal results and wasting money and work power. It should be added that assuring the representativeness of data is most difficult and only rarely achieved in Rn surveys. (One example of a nearly representative Rn survey is the first Austrian Rn survey (6); in the framework of MetroRADON (3), the topic has been addressed in a questionnaire to authorities concerned with Rn (7)). Representativeness is a condition for the accuracy of a result (i.e. no bias or systematic uncertainty). In contrast, the precision of a result is related to the uncertainty of individual data and for aggregates (such as means), and to data dispersion and sample size, that is, number of data.

In planning and performing a survey, one distinguishes between design- and model-based approaches. In the former, the survey is planned such that the target can be directly inferred from the data. For example, if the target is the estimate of the mean indoor Rn concentration in a region with given precision (e.g. in terms of a tolerated confidence interval), a representative sample with minimum size must be generated.

A model-based approach applies models in addition to achieve the target. Geostatistical modeling is common for mapping, which exploits the autocorrelation structure of the quantity to be mapped, or includes auxiliary variables or covariates such as in co-kriging; other methods rely on regression against explanatory or predicting quantities (e.g. indoor Rn explained by uranium concentration in the ground). For this approach, the strict requirements of design-based approaches such as representativeness of the data are relaxed, but model QA is even more demanding, in general (see below).

Data QA

Data QA essentially concerns 1) classical metrological QA for measurement data and critical appraisal of the quality of supportive data such as geological maps. 2) A second, often less emphasized aspect is the QA of experimental protocols and the actual experimenting.

As examples from Rn measurement, aspect 1) deals with traceability to primary standards, proper calibration, reproducibility, and repeatability[2] under controlled laboratory conditions; since there is abundant literature about aspect 1), it is not further discussed here.

On the contrary, aspect 2) deals with proper positioning of a detector in a house and in a room, distance from the wall (for thoron-sensitive detectors), correct exposure time, etc. Also, the correct assignment of metadata belongs to this section, for example, has the floor in which the measurement was performed, truly been the ground floor, as ticked in the questionnaire that accompanies the measurement? However, this latter issue represents a class of sources of uncertainty, which is very difficult to quantify. Altogether, reproducibility and repeatability under largely uncontrolled (and uncontrollable) 'field' conditions are much more difficult to guarantee and to verify than under laboratory conditions.

A particular type of data uncertainty is *ontological and semantic uncertainty*, which occurs if features are defined ambiguously or identified differently by different evaluators (9, 10). A related source of uncertainty is *scale and resolution dissonance* (11), a common problem with the geological and other maps: if a sample point is assigned the geological unit in which it is located, the result may depend on the resolution of the geological map used, since it may be located on a 'small' feature that is not resolved on a coarse map. Similarly, tortuosity of geological borders is increasingly smooth with decreasing resolution, so that a sampling point close to a geological border may, in reality, lie on one side, but is displayed on

---

[2] *Repeatability* quantifies whether an observation or experiment, performed under the same conditions, that is, with the same instrument, by the same person, in the same lab, under the same meteorological conditions, etc., yields the same result (up to statistical tolerance). *Reproducibility* quantifies whether the same result (up to statistical tolerance) is achievable or has been achieved by different methods [Taken from (8), p. 44].

the opposite side. These types of uncertainty are particularly difficult to quantify.

Evaluation QA

Evaluation QA is concerned with adequate statistical methodology, data aggregation, adequate models, mapping methods and their correct application, and uncertainty budgeting.

Adequate and correct models refer to their specification, regarding, for example, assumed type of functional dependence and inclusion of covariates. Further, it pertains to the correct choice of model parameters, such as variogram parameters in geostatistics or parameterization of logistic regression models.

Challenges in mapping methodology include choice of map resolution or pixel size, aggregation or interpolation method, and the choice of continuous or classed level scales of the mapped quantity. Too low resolution may conceal relevant spatial variability, while too high resolution may not be supported by data resolution and suggest precision which cannot be achieved given the data. Aggregation methodology becomes relevant for choropleth maps (spatial units, e.g., municipalities, are assigned a value), while interpolation is the key problem for isopleth maps, where a value is estimated for every point (practically: pixel) of the map.

Establishing an uncertainty budget can be demanding for aggregates of autocorrelated spatial variables and if response variables are modeled from predictors, possibly of different types. At least, one should attempt to identify the sources of uncertainty, which are the components in the uncertainty budget. This may be especially tricky for model uncertainty (an often overlooked component) which encompasses the uncertainty related to the correct choice of a model, its parameterization (because model parameters often result from an estimation procedure), and estimation and prediction of confidence intervals. In many cases, quantification is only feasible by simulation, for example, of bootstrap type, often computationally demanding.

Decision QA

We may distinguish two aspects of decision QA: 1) dealing with competing interests that contribute to decision-making; these are decision predictors whose nature is not the type of data; 2) assessing the chance of ill-decision and uncertainty propagation to a decision from its predictors.

Competing interests

Decisions are based on arguments and interests, which may be competing. Rn policy costs money through surveys, provision against Rn, and Rn remediation. This can be in conflict with the economic interests of private and communal property owners and administrations which are liable to budgetary rigour. Usually, different stakeholders participate in the decision process. While this is fair, QA consists in the transparency of the decision process, meaning that stakeholder positions and the weights given to them should be disclosed.

A particularly sensitive topic is weighing public health against costs. One may argue that in certain cases, remediation is more expensive than the detriment (health cost), or that under the constraint of limited resources, one must prioritize remediation to situations with higher cost-effectiveness. While the conclusion may be correct from a pure cost-effectiveness point of view, its rigid implementation as a kind of *triage*, that is, distinguishing between persons or situations where an intervention has a different chance of success and applying it only to those with a high chance (in the end, monetizing the value of human life), is certainly contrary to what is considered ethically allowed in European culture. Albeit certainly an extreme conclusion, it shows that ethical constraints should be placed in the ‘predictor space’ that forms the set of arguments underlying a decision, without, of course, denying deliberation of cost-effectiveness. As a recent example, in the COVID crisis, most European countries chose to largely ignore budgetary arguments and put all resources available into fighting the pandemic. In Austria, for example, the keyword was until recently ‘*at all costs*’, that is, state budget deficit should not be an argument against the target to keep the health system intact.

A profound analysis of the cost-effectiveness of the Rn abatement policy has been presented in the study by Gray et al. (12). The authors estimate the cost-effectiveness of current (at the time, 2009) and possible alternative policy scenarios for the UK (probably valid in tendency for other countries too). They find that certain alternative scenarios would be more cost-effective, but *do not* draw the conclusion that abatement should be prioritized for certain groups of people at the expense of others, thus – in our opinion – providing a good example of how cost-effectiveness analysis could indeed be used to make Rn policy more efficient.

Concluding this subsection, a *transparent discussion* of how weights are distributed between conflicting arguments that contribute to decisions should be an important part of decision QA.

Probability of erroneous decision, error propagation

A more formal aspect of decision QA is the assessment of the probability of wrong decisions, given data (hard and soft) and methods. Consider two simple real-world examples:

Example 1: Depending on the true value X of a quantity, a decision about action shall be taken. The abstract decision rule ***R*** is: If X > AL (AL represents an action level), action A should be initiated; if $X \leq AL$, no action is necessary (or action B). In practice, a measurement result x of X with uncertainty dx (standard deviation, confidence interval) is compared with an AL. Due to the uncertainty dx, a quantity X measured $x < AL$ can in reality be $>AL$ with some probability. If this probability is high, the abstract decision rule ***R*** does not lead to a reliable decision for a certain range of x – reliable in the sense that the chance of erroneous decision be low.

Decision rules must be specified accordingly; there are the following possibilities, see also the visualization, Fig. 2:

1) 'naive' implementation of the abstract rule ***R***: If $x > AL$, then do A else do B; that is, ignore the uncertainty of x.

To enable reliable decision, modify ***R*** to ***R'***:

2a) If $x > AL-\alpha \times \delta x$, then do A else do B: minimize 2nd kind error, that is, be on the safe side by applying A.

2b) If $x > AL + \beta \times \delta x$, then do A else do B: minimize 1st kind error, that is, avoid unnecessary action due to choosing A.

α and β are set according to the maximum tolerated 2nd or 1st kind errors δA and δB. One may speak of error propagation between decision space (A,δA) and predictor space (x,δx): Uncertainty of x induces chance for erroneous decision, while required reliability of decision acts back to margins (α, β) of predictors.

3) Allow nondecision: If $x > AL + \beta \times \delta x$, then do A because then A is certainly correct; if $x < AL-\alpha \times \delta x$, then do B because then B is certainly correct; if $AL-\alpha \times \delta x \leq x \leq AL + \beta \times \delta x$, then no decision about action is taken, because neither A nor B is reliable; repeat the experiment and collect more data.

Evidently, the choice of decision rule is itself a decision, which may depend on interests in what shall be achieved with priority: action allowing possibly high 1st or 2nd kind errors deemed irrelevant (1), conservative action (2a), parsimonious action (2b) or allowing additional resources (3).

α and β are easily calculated for normally distributed errors dx. For (2a): find x, such that $(1/2)\,\mathrm{erfc}((AL-x)/(s\sqrt{2})) = \delta A$, s – standard deviation of x, erfc – the complementary error function, δA – 2nd kind error; by inversion, $x = AL-\sqrt{2}s\,\mathrm{erfc}^{-1}(2\delta A)$; this shall be equal $AL-\alpha\,\delta x$, hence $\alpha = \sqrt{2}(s/\delta x)\mathrm{erfc}^{-1}(2/\delta A)$. The analogue applies to β in (2b). However, in complex cases, this can be tricky.

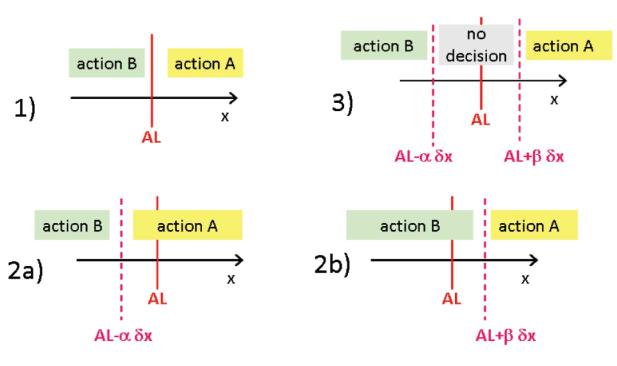

Fig. 2. Visualization of example 1 (see text). Four possible decision rules, according to what shall be optimized.

Example 2: Assume that a municipality is labelled radon priority area (RPA), if in ground floor living rooms of residential buildings the long-term mean indoor Rn concentration (IRC), the RL 300 Bq/m³, is exceeded in more than 10% of cases. The fraction P = prob (IRC > RL) is estimated as p from a sample, usually a small subset of all buildings. Hence p has uncertainty δp, and therefore, by simply comparing p with 10%, a decision about the RPA status of the municipality can be wrong. Like in the first example, 1st and 2nd kind errors are possible, which may have far-reaching consequences, since unnecessarily declaring an area RPA can be expensive (due to action to be implemented, possible legal consequences of ill-assignment), as can be action omitted erroneously (because of the possible detriment that has not been avoided).

It is clear that 1st and 2nd kind errors cannot be avoided. Further, given data or information, they cannot be minimized independently, but there is always a trade-off between them. (Only modifying the experimental design, for example, through denser Rn surveys, that is, larger samples, can reduce them.)

Evidently, the matter can become very complex if the predictor space consists not only of one quantity (X or P in the examples), but of several ones, and even more tricky, if soft data are involved, such as possibly competing economical or political interests (discussed above). While the simple examples were of the type 'if $x < x_0$ then do A else do B', in the real world it could be that 'if $x < x_0$ and if there is enough budget, then do A else do B'. Weighing such 'soft' factors and assigning 'uncertainty' to them seems to be a problem not yet addressed in studies about the implementation of Rn action plans. Yet, perceived ill-weighing can impair the credibility to stakeholders of the resulting decision. Perception depends strongly on the transparency and plausibility of arguments.

Again formally, the decision amounts to *classification*. According to the decision rule, the predictor space can be classified into 'areas' that correspond to 'do action A', 'do action B', etc. In this sense, decision QA could be understood as the attempt to quantify misclassification probability.

Summing up, challenges include these issues:

- Predictors and criteria have uncertainty.
- Predictors of several types (hard, soft; numerical, categorical; etc) contribute to outcomes (estimates, decisions,...) and their uncertainty budgets.
- Identifying the relative weight of predictors, in particular of interests.

- Predictors may not be sufficient for clear decision, leading to 'decision paralysis'.
- Assessing the correctness and reliability of decisions (i.e. chance of wrong decision).

Conclusions

We proposed to understand the quality of a decision (e.g. about measures to reduce Rn exposure) as the quality of the links of the chain which leads to it, encompassing data and their generation, their aggregation into decision predictors, various possibly competing interests that contribute to decisions, and the possibility of erroneous decisions and their consequences.

While metrological QA is well developed and advanced for Rn measurement, design and evaluation QA are much less perceived as important. Decision QA has rarely been discussed at all in the context of Rn policy so far. However, understanding decision processes – here focussing on the Rn abatement policy – is not easy and should be discussed more profoundly, in our opinion, as a condition to be able to assess its quality.

This is important because wrong decisions can be expensive – economically as well as politically. Decisions must be legally proven and defendable, which requires that the pathway leading to them must be quality assured. Likewise, credibility to and acceptance by stakeholders – the public in the first place – depend on their proven soundness.

Conflict of interest and funding

The authors declare no conflict of interest. This article is partly based on previous work that was supported by the European Metrology Programme for Innovation and Research (EMPIR), JRP-Contract 16ENV10 MetroRADON (www.euramet.com). The EMPIR initiative was cofunded by the European Union's Horizon 2020 research and innovation programme and the EMPIR Participating States.

References

1. Zeeb H, Shannoun F, eds. WHO handbook on indoor radon: A public health perspective. 2009. Available from: www.who.int/publications/i/item/9789241547673 [cited 21 June 2021].
2. European Commission (EC). Council Directive 2013/59/Euratom of 5 December 2013 laying down basic safety standards for protection against the dangers arising from exposure to ionising radiation, etc. Official Journal of the European Union 2014; 57(L13): 1–73. Available from: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L:2014:013:FULL&from=EN [cited 21 June 2021].
3. MetroRADON. Available from: http://metroradon.eu/ [cited 21 June 2021].
4. Guo, K L (2008) DECIDE. The Health Care Manager, 27(2), 118–127. doi: 10.1097/01.hcm.0000285046.27290.90; https://en.wikipedia.org/wiki/Decision-making#DECIDE (accessed 20 June 2021)
5. Deleuze G, Guattari F. Capitalisme et Schizophrénie 2 – Mille Plateaux. Paris: Minuit; 1980. ISBN 978-2707303073. Available from: https://fr.wikipedia.org/wiki/Rhizome_(philosophie). [cited 10 March 2022]
6. Friedmann H. Final results of the Austrian radon project. Health Phys 2005; 89(4): 339–48. doi: 10.1097/01.hp.0000167228.18113.27
7. Cinelli G, Bochicchio F, Bossew P, Carpentieri C, De Cort M, Gruber V, et al. Similarities and differences between radon surveys across Europe: Results from MetroRADON questionnaire. JERA 2021. Journal of the European Radon Association 2022, 3: 7605. doi: 10.35815/radon.v3.7605
8. European Commission (EC), Joint Research Centre, Cinelli G, De Cort M, Tollefsen T, eds. European atlas of natural radiation. Luxembourg: Publication Office of the European Union; 2019. Printed version: ISBN 978-92-76-08259-0. doi: 10.2760/520053; Catalogue number KJ-02-19-425-EN-C; Online version: ISBN 978-92-76-08258-3; doi: 10.2760/46388; Catalogue number KJ-02-19-425-EN-N. Available from: https://remon.jrc.ec.europa.eu/About/Atlas-of-Natural-Radiation/Download-page [cited 24 June 2021].
9. Fisher P. (2008) Uncertainty, Semantic. In: Shekhar S., Xiong H. (eds) Encyclopedia of GIS. Springer, Boston, MA. doi: 10.1007/978-0-387-35973-1_1425
10. Fox S. Ontological uncertainty and semantic uncertainty in global network organizations. 2008. Available from: www.vttresearch.com/sites/default/files/pdf/workingpapers/2008/W102.pdf [cited 22 June 2021].
11. Bossew P, Čeliković I, Cinelli G, Ciotoli GC, Domingos F, Gruber V, et al. On harmonization of radon maps. JERA 2021. Journal of the European Radon Association 2022, 3: 7554. doi: 10.35815/radon.v3.7554.
12. Gray A, Read S, McGale P, Darby S. Lung cancer deaths from indoor radon and the cost effectiveness and potential of policies to reduce them. BMJ 2009; 338: a3110. doi: 10.1136/bmj.a3110

***Peter Bossew**
German Federal Office for Radiation Protection Berlin
Germany Köpenicker Allee 120-130, 10318 Berlin.
Email: pbossew@bfs.de