MEMORANDUM

|  |  |
| --- | --- |
| **To:** | Internet Corporation for Assigned Names and Numbers, EPDP Team |
| **From:** | Ruth Boardman & Phil Bradley-Schmieg |
| **Date:** | 9 April 2021 |
| **Subject:** | March 2021 question regarding options for contact address masking |
|  |  |

Background

The European Data Protection Board (“EDPB”), in [a July 2018 letter to Göran Marby](https://www.icann.org/en/system/files/correspondence/jelinek-to-marby-05jul18-en.pdf), stated that:

“*personal data identifying individual employees (or third parties) acting on behalf of the registrant should not be made publicly available by default in the context of WHOIS*”.

Against this background and building on previous advice you have received in this matter, you have raised the following question.

|  |
| --- |
| ***Question presented:*** B&B’s [Memo dated 4 February 2020 regarding email contact information](https://community.icann.org/download/attachments/111388744/Memo%20-%20ICANN%20-%2004.02.2020.docx?version=1&modificationDate=1581360214000&api=v2) discussed two options: (a) a “pseudonymous email contact” where the same unique string is used for multiple registrations by the data subject; and (b) an “anonymous email contact” where a separate unique email string is used for each such registration.  B&B opined that publication of either (a) or (b) would be treated as publication of personal data on the web because the purpose of making this masked email address available is to allow 3rd parties to directly contact the data subject and because third parties with legitimate and proportionate interests would have access to the underlying data.Upon review, the EPDP Legal Team has proposed to describe options (a) and (b) going forward as follows:* The phrase "pseudonymous email contact” (option (a)) should be replaced with the phrase "**Registrant-based email contact**," defined as: “an email for all domains registered by a unique registrant, *which is intended to be pseudonymous data when processed by third party users* (i.e., non-contracted parties). (The question of whether the email should be common across ICANN-accredited Registrars requires a policy determination TBD.)
* The phrase "anonymous email contact" (option (b)) should be replaced with the phrase "**Registration-based email contact**," defined as “a separate single use email for each domain name registered by a unique registrant, *which is intended to be virtually or “essentially” anonymous data when processed by third party users* (i.e., non-contracted parties).”

In answering the questions below, please assume, for discussion purposes, that third-party users of Registration-based email contact information cannot identify the data subject without disproportionate effort so that the risk of identification appears in reality to be insignificant. 1. Based on your experience and applicable precedent, please compare the level of risk, likelihood of enforcement actions, fines, counseling, etc. associated with (a) publication on the web or (b) automated disclosure of (i) a Registrant-based email contact on the one hand and (ii) a Registration-based email contact on the other?  In responding to this question please consider:
	1. Whether the assumed fact that the risk of data subject identification by a third party (i.e., non-contracted party) through a Registration-based email contact appears to be insignificant would render such emails effectively “anonymous” with respect to such third parties under the *Breyer*standard?
	2. If not, how would the choice of email contact (Registrant-based or Registration-based) affect the outcome of the legitimate interests balancing test under Article 6(1)(f)? To what extent would the use of a Registration-based email contact reduce the impact of publication on the interests or fundamental rights and freedoms of the data subject?

Does the answer to these questions change if the primary purpose for publishing a masked email is to support statistical research and analytics, and not to communicate with the data subject? |

*Analysis*

Our answer starts by addressing your sub-question, “*Whether the assumed fact that the risk of data subject identification by a third party (i.e., non-contracted party) through a Registration-based email contact appears to be insignificant would render such emails effectively “anonymous” with respect to such third parties under the Breyer standard?*”, to explain why we consider that the GDPR would remain applicable in a Registration-based email contact scenario. We then turn to the wider GDPR compliance aspects of your question.

***Anonymity***

We maintain our view, expressed in our Memorandum dated 4th February 2020, that with either option (Registrant-based or Registration-based email contact), there remains a high likelihood that the publication or automated disclosure of such email addresses would be considered to be the processing of personal data.

For the GDPR to apply to the processing of electronic data (assuming the GDPR’s territoriality test is met, and its subject matter carve-outs are not applicable), a two-part test applies:

* 1. First, there must be processing of information that relates to a *particular* individual, having regard to the data (and its processing’s) “content, purpose, or effect”. This is the “*Nowak”[[1]](#footnote-1)* / “relates to” test.
	2. Second, that particular individual must be “identified or identifiable”, which means that there must exist “means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly.”[[2]](#footnote-2) “Identification” does not necessarily mean finding the real name of a person; rather, it has a more general meaning, generally revolving around the ability to specifically “single out” someone for different treatment (singling out), [[3]](#footnote-3) and/or having the ability to collect/connect more data about them (inference and/or linking).[[4]](#footnote-4) A technical identifier – even one that was randomly generated – can be sufficient for such purposes, particularly if it is linked with other information about the person that makes it easier to distinguish them from someone else.[[5]](#footnote-5) There are no “reasonably likely means” of reidentification if such activity is “prohibited by law or practically impossible on account of the fact that it requires a disproportionate effort in terms of time, cost and man-power, so that the risk of identification appears in reality to be insignificant”[[6]](#footnote-6). This is the “*Breyer*” / “identifiability” test.

Our view, expressed above, is that the processing of these email aliases would still likely be seen as meeting both tests, to the extent that the purpose of the processing is to provide a means of contacting data subjects.

*Nowak* test

Regarding the *Nowak* test: when a contact is a natural person, such addresses will be masked aliases for a real email address used by that person. In light of this:

* 1. Where the purpose / intended effect of the processing of that data is to enable correspondence with the recipient (i.e., often, with a specific data subject), then having regard to the EU Court of Justice (“CJEU”)’s test in *Nowak*, that “purpose” and/or “effect” means there is a link to a *particular* individual.[[7]](#footnote-7)
	2. By contrast, purely statistical processing aimed at creating *aggregate* metrics (describing relatively large cohorts) – e.g. counting how many such contact aliases have been created – may arguably *not* be subject to the GDPR. This is because the *content* of a randomly-generated contact alias does not specifically link to a specific individual, at least in a Registration-based email contact scenario; and – again, arguably – neither the *purpose* nor the *effect* of creating aggregate results of statistical research carries a link to a *particular* individual; rather, aggregate statistics describe and differentiate between *cohorts/groups* (e.g. by nation, Registry, Registrar, etc.). The *Nowak* test may arguably not be satisfied in respect of that class of processing (but note that this is to be distinguished from statistics aimed at generating new information about, or classification of, any specific data subject – e.g. counting how many domain names are associated with a given Registrant-based email contact).
	3. However, in practice we do not think it would be reasonably possible to say that the sole purpose of creating and publishing the contact aliases is for the aggregate statistical processing just described. If this were the case, there would be no need to provide an email address at all. The fact that an email address is provided suggests that a significant purpose for the creation and publication of contact aliases will always be to provide a means of contacting specific persons. Accordingly, while *some* processing (for aggregate statistics) may fall outside the GDPR’s scope based on the *Nowak* test, the GDPR seems likely to remain a compliance concern at the very least in respect of the *other* purpose of processing.
	4. We should also caution against over-reliance on *Nowak-*based arguments. Despite the ruling echoing early Article 29 Working Party guidance,[[8]](#footnote-8) we are not aware of the *Nowak* test being systematically applied in the analyses and guidance of courts and supervisory authorities applying the GDPR. For example, as of early April 2021, a search of the Belgian Data Protection Authority’s website, across all available languages, turns up (i) just two directly references to the *Nowak* case, and only on unrelated points; and (ii) apparently no citations of the key “content, purpose or effect” phrase from *Nowak*. That authority’s explanation (in its Lexicon) of the term “personal data” concentrates exclusively on the *Breyer* test – i.e. identifiability of a data subject.[[9]](#footnote-9) Other authorities may take a different view (e.g. the UK authority does discuss the “content, purpose or effect” test, and summarises its impact as follows: “Information must ‘relate to’ the identifiable individual to be personal data. This means that it does more than simply identifying them – it must concern the individual in some way. (…) Data can reference an identifiable individual and not be personal data about that individual, as the information does not relate to them.”)[[10]](#footnote-10)
	5. Moreover, not only do authorities in this field not always place substantial emphasis on *Nowak*, but *if* they were do so, they could also take quite differing approaches to its interpretation. Differences of opinion might in particular surround the “content” limb of the “content, purpose or effect” test. Article 29 Working Party, *Opinion 4/2007 on the concept of personal data* (WP 136)[[11]](#footnote-11) explained that “[t]he “content” element is present in those cases where - corresponding to the most obvious and common understanding in a society of the word "relate" - information is given about a particular person, regardless of any purpose on the side of the data controller or of a third party, or the impact of that information on the data subject.” If that explanation is correct, then a court or regulator might conclude that publishing an email address (even a randomly generated one) for a contact associated with a domain registration is *inherently* publishing information “about” that person – because it tells us how to contact that person. This is a problematic view, however, as it “borrows” reasoning from the *purpose* and *effect* tests (it looks at a possible purpose for the information, *not* at the content of the information itself), *and* bases itself on a *hypothetical* purpose/effect, not the *actual* purpose/effect of processing – thus completely short-circuiting two thirds of the “content, purpose or effect” test. From both a logical and rule of law (clarity/certainty) perspective, this is problematic. From a simpler point of view, something randomly generated (as876bnk@example.com) is a pure expression of random “noise” -- an instantaneous snapshot of the electrical state of a computer’s “random number generator” circuitry. It thus does not and cannot of itself “*contain*” any information about any person. If it *did* in and of itself convey information about a person, it logically would not be random. From that view, a randomly-generated address thus does not pass the “content” test; instead, the focus would need to be on the data processing’s purpose and/or effect.
	6. Clearly, then, there is a significant risk of disagreement with at least some authorities if arguments rest on the *Nowak* case.

*Breyer* test

Regarding the *Breyer* test: in that case, the CJEU constructed a thought experiment: if there was a cyber attack, a controller holding an IP address (and, we presume – though the court is not explicit on this point – a timestamp indicating when that IP address was in use by a device/person of interest), could communicate that information to the police/judicial authorities. The CJEU expected that the authorities would then often be empowered to then demand corresponding information from the internet access provider that assigned that IP address, and thereby bring a prosecution (although the CJEU asked the referring national courts to verify that assumption). The CJEU thus held that unless this scenario was prohibited by law or practically impossible, there were “reasonably likely means” of identifying a data subject.

The key point here is that although a third party may just know a Registrant-based or Registration-based email contact, competent authorities could correlate this to non-public registration data held by Contracted Parties, allowing for reidentification. So far as we are aware, this would not always require “practically impossible” levels of effort, nor would it be universally prohibited by law.

Thus even from the perspective *of third parties*, the distribution and use of such contact aliases could be treated as personal data processing.

From the perspective *of a Contracted Party* that knowswhich contact alias it has assigned to a Registrant / Registrant’s nominated contact, the creation and hosting of such addresses, and their making available for use by others, is almost certainly personal data processing (when the contact persons are natural persons).

*Risk of the respective options presented*

Having explained our view that for either option, the GDPR remains relevant, we turn now to your request that we compare risks associated with (a) publication on the web or (b) automated disclosure of (i) a Registrant-based email contact on the one hand and (ii) a Registration-based email contact on the other.

Our summary (which reflects the important assumptions and caveats provided later in this answer) is as follows:

|  |  |  |
| --- | --- | --- |
|  | Registrant-based email contact | Registration-based email contact |
| Web publication | Medium | Low |
| Automated disclosure | Low | Lowest |

Based on an application of the GDPR’s principles, the sharing (whether through web publication or automated disclosure) of Registration-based email aliases carries lower risk compared to Registrant-based email aliases.

This is because someone holding a Registrant-based email address may be able to learn more information about the data subject – specifically, what other domain names that data subject is associated with. This is because unless a different *real* contact address was provided for that data subject for each domain they register, then each registration would carry the same email alias.

Web publication of such details could make it relatively easy to build such profiles and potentially even build a reverse lookup tool (‘for a given Registration-based email contact, what domain names is this contact associated with?’).

Automated disclosure, alone, would presumably make this more difficult, since unless the automated disclosure tools *specifically* provide reverse lookup functionality,[[12]](#footnote-12) requesters would presumably need to query potentially quite large numbers of domain names to gather enough information to be able to make matches and start to build an (incomplete) reverse lookup function. That said, requestors that have a pre-established list of specific domain names (e.g. suspected “mirrors” of a website hosting illegal contents) could determine whether the same email address was provided for some or all of those sites. Thus even in an automated disclosure scenario, the use of a Registrant-based email contact scheme carries added risks to privacy, relative to Registration-based email contact scheme.

Accordingly, having regard to the following considerations:

* 1. The need to comply with the GDPR’s data minimisation rule;
	2. The need to comply with a “privacy by design and by default” rule;
	3. That reliance on GDPR Article 6(1)(f) (the legitimate interests legal basis) is more robust when system design minimises prejudice to “the interests or fundamental rights and freedoms of the data subject which require protection of personal data”; and
	4. That in assessing whether and to what extent fines should be levelled against a controller, authorities must have regard *inter alia* to the “gravity” of an infringement, the “scope” of processing, the “the level of damage suffered by” data subjects, “any action taken by the controller or processor to mitigate the damage suffered by data subjects” and “the degree of responsibility of the controller or processor taking into account technical and organisational measures implemented by them pursuant to Articles 25 and 32” (see GDPR Article 83),

we therefore consider that a Registration-based email contact scheme carries lower risk than a Registrant-based email contact scheme.

Having explained the balance of risk along the “Registration vs. Registrant-based scheme” axis, we turn now to contrasting risks for web-based publication versus automated disclosure.

A risk common to both a Registration-based and Registrant-based email contact schemes is spam or other unsolicited emails; this “addressability” is, arguably, one aspect of privacy.[[13]](#footnote-13) Spam has been a longstanding concern for WHOIS systems; it was the subject of an ICANN Security and Stability Advisory Committee study in 2007, which concluded that “the appearance of email addresses in response to WHOIS queries is indeed a contributor to the receipt of spam, albeit just one of many”.[[14]](#footnote-14)

Accordingly, whether a Registrant- or Registration-based email contact system is employed, effective measures should be taken to address the availability of addresses to spammers (e.g. use of technical features to prevent “harvesting” of such addresses; and/or filtering out inappropriate communications before they are delivered to the intended recipient).

In comparison to web-based publication, we presume that automated disclosure allows further scope to evaluate the motives for a request, the sources of that request, and to monitor / audit and apply protective measure (e.g. rate limits) on such requests – i.e. greater scope to deploy the sorts of mitigations that will reduce liability based on the factors set out in paragraph 18 above. It would therefore appear that automated disclosure poses inherently less risk on this front, compared to web-based publication.

Those potential advantages of automated disclosure compared to web-based publication also conceivably present GDPR Article 25 (privacy by design and by default) advantages. Particularly, some thought would need to be given to ensuring that web-based publication is designed in such a way that it complies with GDPR Article 25(2), “such measures shall ensure that by default personal data are not made accessible without the individual's intervention to an indefinite number of natural persons”.[[15]](#footnote-15)

That said, if effective measures against spam are employed, and if a Registration—based approach is taken (due to its advantages discussed earlier), then given the resulting low utility of the data, it is difficult to see how its web-based publication would present meaningful risks to privacy or data security.

\* \* \*

1. Judgement of the CJEU in Case C‑434/16 *Nowak,* ECLI:EU:C:2017:994, at paragraph 35. [↑](#footnote-ref-1)
2. GDPR Recital 26 [↑](#footnote-ref-2)
3. As quoted above, GDPR Recital 26 specifically refers to “singling out” when discussing means that are reasonably likely to be used to identify the data subject. [↑](#footnote-ref-3)
4. Singling out, linkability and inference are three parts of the anonymisation test proposed by the Article 29 Working Party, in its Opinion 05/2014 on Anonymisation Techniques (“WP 216”), available online at <https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf> [↑](#footnote-ref-4)
5. On this point, see GDPR Recital 30 (“Natural persons may be associated with online identifiers provided by their devices, applications, tools and protocols, such as internet protocol addresses, cookie identifiers or other identifiers such as radio frequency identification tags. This may leave traces which, in particular when combined with unique identifiers and other information received by the servers, may be used to create profiles of the natural persons and identify them.”) [↑](#footnote-ref-5)
6. Judgement of the CJEU in Case C‑582/14 *Breyer,* ECLI:EU:C:2016:779, at paragraphs 45 and 46. [↑](#footnote-ref-6)
7. In some cases, a recipient contact address might be a shared mailbox (e.g. enquiries@example.com), in which case the masked contact address is arguably *not* personal data, whether by application of the *Nowak* or *Breyer* tests. [↑](#footnote-ref-7)
8. WP 136, at page 10. [↑](#footnote-ref-8)
9. <https://www.autoriteprotectiondonnees.be/citoyen/vie-privee/lexique> [↑](#footnote-ref-9)
10. <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/what-is-personal-data/what-is-the-meaning-of-relates-to/#pd5> [↑](#footnote-ref-10)
11. Article 29 Working Party, *Opinion 4/2007 on the concept of personal data* (WP 136), at p. 10. Available online at <https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf> [↑](#footnote-ref-11)
12. Such features, before being rolled out, would require careful consideration. For old guidance on the issue, see Article 29 Working Party Opinion 5/2000 on The Use of Public Directories for Reverse or Multi-criteria Searching Services (Reverse Directories) (“WP 33”), available online at <https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2000/wp33_en.pdf> [↑](#footnote-ref-12)
13. Recital 40 of Directive 2002/58/EC (the EU’s “ePrivacy Directive”) states: “Safeguards should be provided for subscribers against intrusion of their privacy by unsolicited communications for direct marketing purposes in particular by means of automated calling machines, telefaxes, and e-mails, including SMS messages.” [↑](#footnote-ref-13)
14. *SAC 023: Is the WHOIS Service a Source for email Addresses for Spammers?* , Executive Summary. Available online at <https://www.icann.org/en/system/files/files/sac-023-en.pdf> [↑](#footnote-ref-14)
15. In *its Guidelines 4/2019 on Article 25 Data Protection by Design and by Default, v2.0*, at paragraph 56, the EDPB explains that this means that “[t]he controller shall by default limit accessibility and give the data subject the possibility to intervene before publishing or otherwise making available personal data about the data subject to an indefinite number of natural persons”. Available online at <https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_201904_dataprotection_by_design_and_by_default_v2.0_en.pdf> [↑](#footnote-ref-15)