Fraunhofer
IESE

CrowdRE, User Feedback and GDPR

Towards Tackling GDPR Implications with Adequate Technical and Organizational Measures in an Effort-Minimal Way

Eduard C. Groen, Michael Ochs
Fraunhofer IESE

Jeju Island
24 September 2019

General Data Protection Regulation

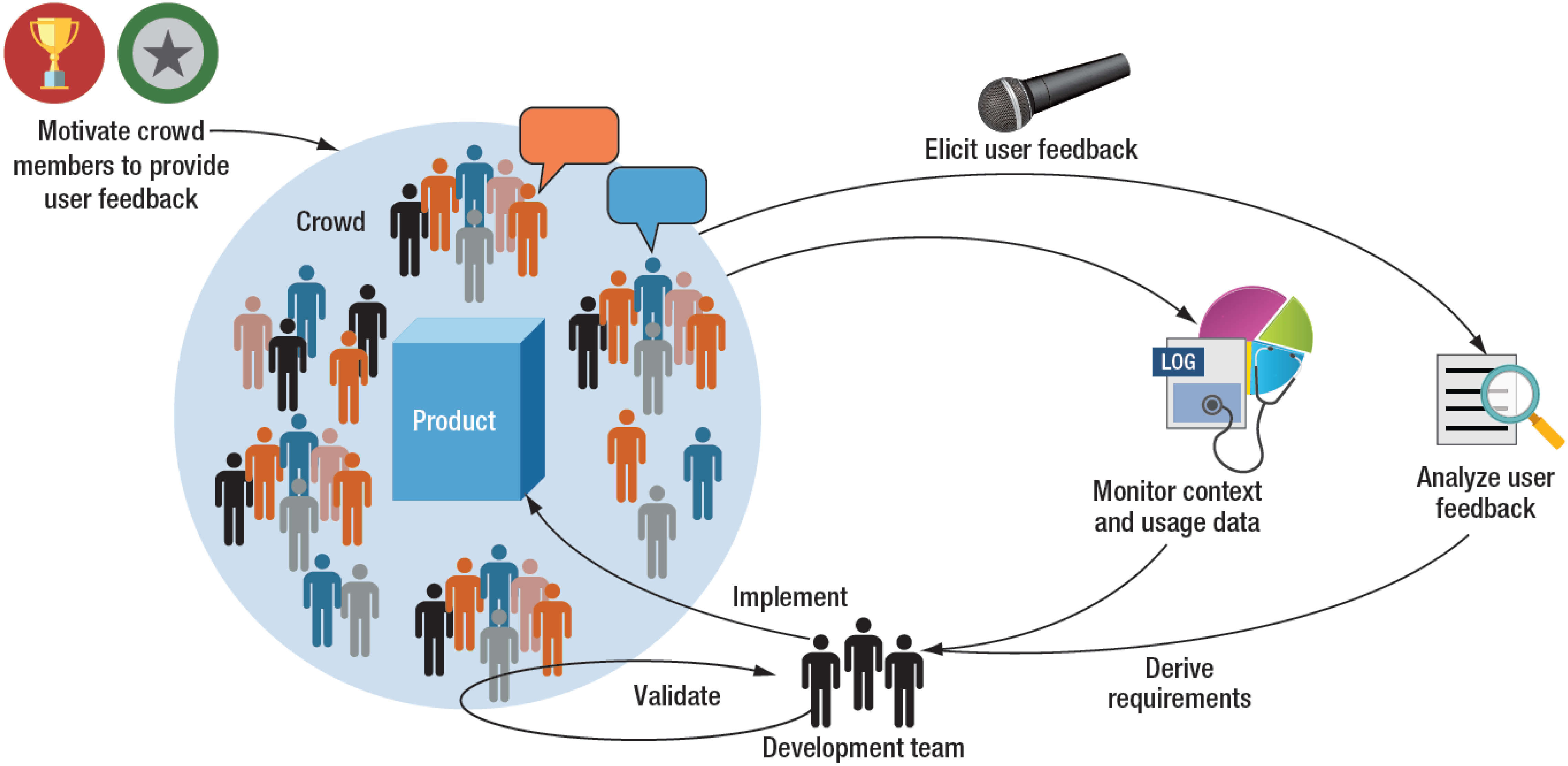
Image source: https://giphy.com/gifs/jaws-gifs-BjASvVZLMWc36

How Not to Respond: Confused, Anxious, Ignorant, Reckless

Image sources:

https://giphy.com/gifs/feeling-covfefe-3jN3GziOKUEmI
https://tenor.com/view/head-in-the-sand-gif-12598832
https://giphy.com/gifs/scared-nervous-ren-and-stimpy-y9X0F8VgTkmU8
https://tenor.com/view/fell-off-car-crash-water-gif-9809532

Does GDPR Affect Crowd-based Requirements Engineering (CrowdRE)?



Source: Groen, E. C., Seyff, N., Ali, R., Dalpiaz, F., Doerr, J., Guzman, E., Hosseini, M., Marco, J., Oriol, M., Perini, A., & Stade, M. (2017). The crowd in requirements engineering: The landscape and challenges. IEEE Software, March/April 2017. Image © 2017 IEEE Society.

Public Sources of Text-based User Feedback

Non-exhaustive

Google Play
Apple App Store
Windows Store
Amazon

YouTube
TestFreaks
Yelp
Sitejabber
SourceForge
Groupon

Redmine
Bugzilla
Jira
Salesforce
SAP CRM
NetSuite

Facebook
Twitter
phpBB
Wordpress
LinkedIn

amazing ★★★★★

by Kelly Strathlyle

Finally an app that is capable of canceling ambient noise!

Positive statement about the product **functionality** “noise cancellation”

ghastly ★☆☆☆☆

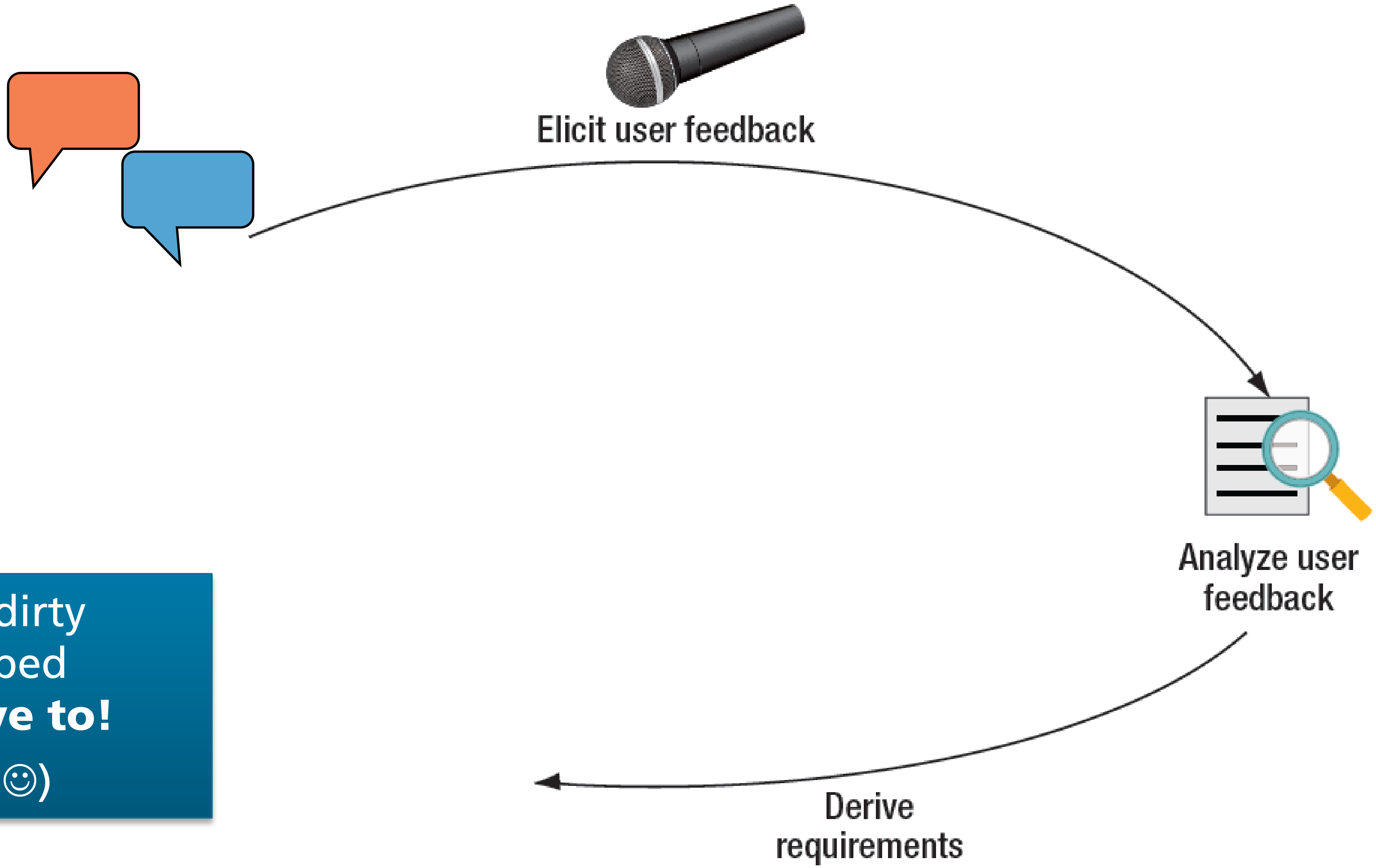
by Kelly Strathlyle

Slow app, clunky interface, and interrupts the music frequently, telling you to buy the pro version. What a ripoff! @strathlyle

Negative statement about the product **quality** “Performance Efficiency”

Image source: FormatF Productions, used with permission

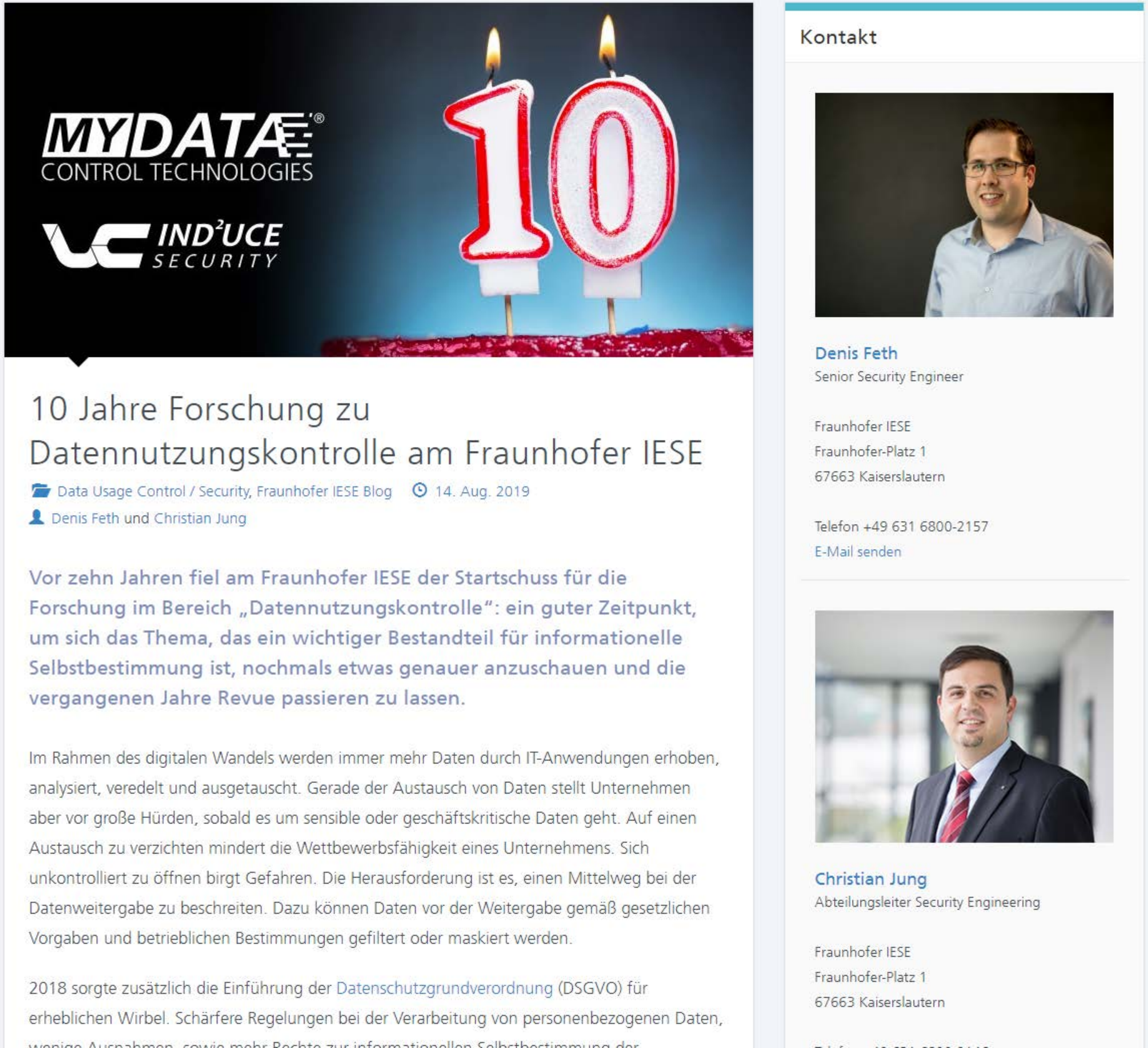
Areas of CrowdRE Potentially Affected by GDPR



We got our hands dirty
and minds numbed
so you don‘t have to!
(You‘re welcome ☺)

Source: Groen, E. C., Seyff, N., Ali, R., Dalpiaz, F., Doerr, J., Guzman, E., Hosseini, M., Marco, J., Oriol, M., Perini, A., & Stade, M. (2017). The crowd in requirements engineering: The landscape and challenges. IEEE Software, March/April 2017. Image © 2017 IEEE Society.

Who are we to talk? 1/2



10 Jahre Forschung zu Datennutzungskontrolle am Fraunhofer IESE

Data Usage Control / Security, Fraunhofer IESE Blog 14. Aug. 2019

Denis Feth und Christian Jung

Vor zehn Jahren fiel am Fraunhofer IESE der Startschuss für die Forschung im Bereich „Datennutzungskontrolle“: ein guter Zeitpunkt, um sich das Thema, das ein wichtiger Bestandteil für informationelle Selbstbestimmung ist, nochmals etwas genauer anzuschauen und die vergangenen Jahre Revue passieren zu lassen.

Im Rahmen des digitalen Wandels werden immer mehr Daten durch IT-Anwendungen erhoben, analysiert, veredelt und ausgetauscht. Gerade der Austausch von Daten stellt Unternehmen aber vor große Hürden, sobald es um sensible oder geschäftskritische Daten geht. Auf einen Austausch zu verzichten mindert die Wettbewerbsfähigkeit eines Unternehmens. Sich unkontrolliert zu öffnen birgt Gefahren. Die Herausforderung ist es, einen Mittelweg bei der Datenweitergabe zu beschreiten. Dazu können Daten vor der Weitergabe gemäß gesetzlichen Vorgaben und betrieblichen Bestimmungen gefiltert oder maskiert werden.

2018 sorgte zusätzlich die Einführung der Datenschutzgrundverordnung (DSGVO) für erheblichen Wirbel. Schärfere Regelungen bei der Verarbeitung von personenbezogenen Daten, wenige Ausnahmen, sowie mehr Rechte zur informationellen Selbstbestimmung der

Kontakt

Denis Feth
Senior Security Engineer

Fraunhofer IESE
Fraunhofer-Platz 1
67663 Kaiserslautern

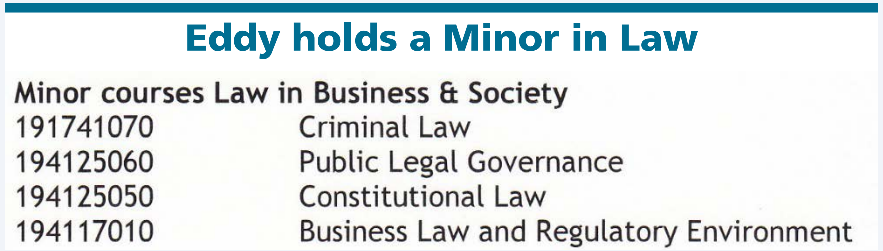
Telefon +49 631 6800-2157
E-Mail senden

Christian Jung
Abteilungsleiter Security Engineering

Fraunhofer IESE
Fraunhofer-Platz 1
67663 Kaiserslautern

https://blog.iese.fraunhofer.de/

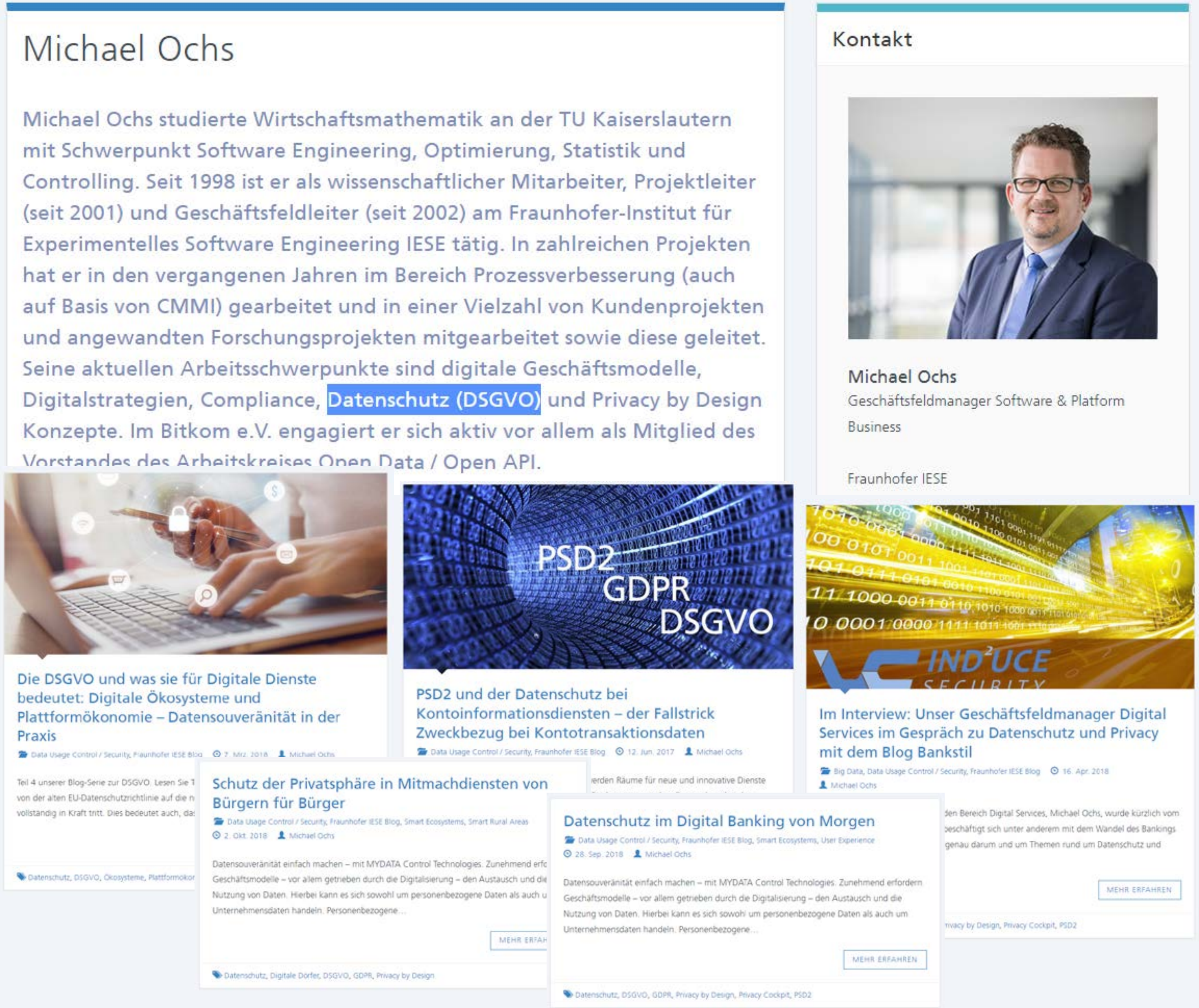
Who are we to talk? 2/2



Eddy holds a Minor in Law

Minor courses Law in Business & Society

191741070	Criminal Law
194125060	Public Legal Governance
194125050	Constitutional Law
194117010	Business Law and Regulatory Environment



Michael Ochs

Michael Ochs studierte Wirtschaftsmathematik an der TU Kaiserslautern mit Schwerpunkt Software Engineering, Optimierung, Statistik und Controlling. Seit 1998 ist er als wissenschaftlicher Mitarbeiter, Projektleiter (seit 2001) und Geschäftsfeldleiter (seit 2002) am Fraunhofer-Institut für Experimentelles Software Engineering IESE tätig. In zahlreichen Projekten hat er in den vergangenen Jahren im Bereich Prozessverbesserung (auch auf Basis von CMMI) gearbeitet und in einer Vielzahl von Kundenprojekten und angewandten Forschungsprojekten mitgearbeitet sowie diese geleitet. Seine aktuellen Arbeitsschwerpunkte sind digitale Geschäftsmodelle, Digitalstrategien, Compliance, Datenschutz (DSGVO) und Privacy by Design Konzepte. Im Bitkom e.V. engagiert er sich aktiv vor allem als Mitglied des Vorstandes des Arbeitskreises Open Data / Open API.

Kontakt

Michael Ochs
Geschäftsfeldmanager Software & Platform Business

Fraunhofer IESE

Die DSGVO und was sie für Digitale Dienste bedeutet: Digitale Ökosysteme und Plattformökonomie – Datensouveränität in der Praxis

Data Usage Control / Security, Fraunhofer IESE Blog · 7. Mrz. 2018 · Michael Ochs

Teil 4 unserer Blog-Serie zur DSGVO. Lesen Sie T
von der alten EU-Datenschutzrichtlinie auf die n
vollständig in Kraft tritt. Dies bedeutet auch, das

Datenschutz, DSGVO, Ökosysteme, Plattformökor

PSD2 und der Datenschutz bei Kontoinformationsdiensten – der Fallstrick Zweckbezug bei Kontotransaktionsdaten

Data Usage Control / Security, Fraunhofer IESE Blog · 12. Jun. 2017 · Michael Ochs

erden Räume für neue und innovative Dienste

Im Interview: Unser Geschäftsfeldmanager Digital Services im Gespräch zu Datenschutz und Privacy mit dem Blog Bankstil

Big Data, Data Usage Control / Security, Fraunhofer IESE Blog · 16. Apr. 2018 · Michael Ochs

den Bereich Digital Services, Michael Ochs, wurde kürzlich vom
beschäftigt sich unter anderem mit dem Wandel des Bankings
genau darum und um Themen rund um Datenschutz und

MEHR ERFAHREN

rivacy by Design, Privacy Cockpit, PSD2

Schutz der Privatsphäre in Mitmachdiensten von Bürgern für Bürger

Data Usage Control / Security, Fraunhofer IESE Blog, Smart Ecosystems, Smart Rural Areas · 2. Okt. 2018 · Michael Ochs

Datensouveränität einfach machen – mit MYDATA Control Technologies. Zunehmend erfo
Geschäftsmodelle – vor allem getrieben durch die Digitalisierung – den Austausch und die
Nutzung von Daten. Hierbei kann es sich sowohl um personenbezogene Daten als auch u
Unternehmensdaten handeln. Personenbezogene…

MEHR ERFAH

Datenschutz, Digitale Dorfer, DSGVO, GDPR, Privacy by Design

Datenschutz im Digital Banking von Morgen

Data Usage Control / Security, Fraunhofer IESE Blog, Smart Ecosystems, User Experience · 28. Sep. 2018 · Michael Ochs

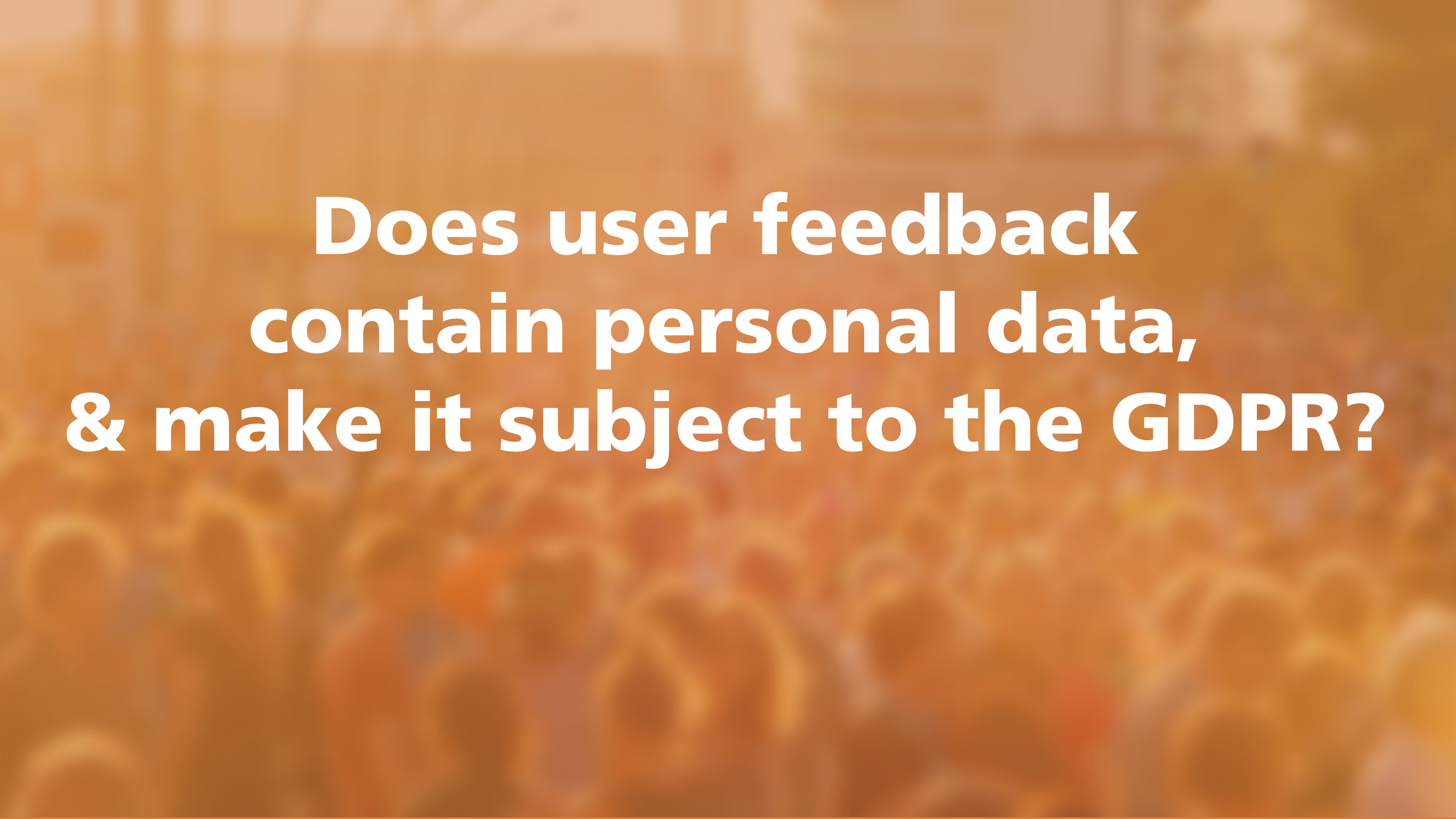
Datensouveränität einfach machen – mit MYDATA Control Technologies. Zunehmend erfordern Geschäftsmodelle – vor allem getrieben durch die Digitalisierung – den Austausch und die Nutzung von Daten. Hierbei kann es sich sowohl um personenbezogene Daten als auch um Unternehmensdaten handeln. Personenbezogene…

MEHR ERFAHREN

Datenschutz, DSGVO, GDPR, Privacy by Design, Privacy Cockpit, PSD2

No legal counsel!

https://blog.iese.fraunhofer.de/



Does user feedback contain personal data, & make it subject to the GDPR?

What is “Personal Data”?

- **Personal data**
  - “any information relating to an identified or identifiable natural person” (Art. 4.1 (1) GDPR)

- **Identifiable natural person (a.k.a. “data subject”)**
  - “one who can be identified, directly or indirectly, in particular by reference to an identifier […] or to one or more factors specific to […] that natural person”

- **Examples:**

Contact Details	Individual Information	Identification Numbers	Personal Views	Health-related Information	Nonlinguistic Information
•Name •Email Address •Home Address •Phone Number •Social Media ID	•Date of Birth •Location Data (e.g., GPS Position) •Ethnic Origin	•Account Number (e.g., IBAN) •Credit Card Number •License Plate Number •Passport Number	•Political Convictions •Religious Beliefs •Philosophical Views	•Blood Pressure •Heart Rate	•Photos •Handwriting •Digital IDs (e.g., IP Address)

amazing ★★★★★

by Kelly Strathlyle

Finally an app that is capable of canceling ambient noise!

ghastly ★☆☆☆☆

by Kelly Strathlyle

Slow app, clunky interface, and interrupts the music frequently, telling you to buy the pro version. What a ripoff @strathlyle

Image source: FormatF Productions, used with permission

Yes, user feedback contains personal data.

Ethical issue: assuring the users’ privacy & anonymity. *

* F. Fotrousi, N. Seyff, J. Börstler, “Ethical considerations on research on user feedback,” in Proc. IEEE 25th Int. Req. Engg. Conf. Workshops, 2017, pp. 194–198.

Which Personal Data Exactly?

- **Quick experiment:**

Contact Details	Individual Information	~~Identification Numbers~~	~~Personal Views~~	~~Health-related Information~~	~~Nonlinguistic Information~~
•Name •Email Address •Home Address •Phone Number •Social Media ID	•Date of Birth •Location Data (e.g., GPS Position) •Ethnic Origin	~~•Account Number (e.g., IBAN)~~ ~~•Credit Card Number~~ ~~•License Plate Number~~ ~~•Passport Number~~	~~•Political Opinions~~ ~~•Religious Beliefs~~ ~~•Philosophical Views~~	~~•Blood Pressure~~ ~~•Heart Rate~~	~~•Photos~~ ~~•Handwriting~~ ~~•Digital IDs (e.g., IP Address)~~
Likely	Maybe	We have never encountered this	Typically no discussions in app stores	*Maybe* found for medical apps at best	We focus on text-based user feedback

- Analysis of dataset by Groen et al. (2017)

Dataset: Groen, E. C., Kopczyńska, S., Hauer, M. P., Krafft, T. D., & Doerr, J. (2017). Users – The hidden software product quality experts?: A study on how app users report quality aspects in online reviews. In: Proceedings of the IEEE 25th International Requirements Engineering Conference, 80–89.

Findings 1/3

- **Date of Birth**
  - *Search strategy:* various date notation formats
  - *Result:* none found, the only mentions of dates were recent, e.g.,
    - a date a review was updated;
    - a report of an incorrect date representation by an app

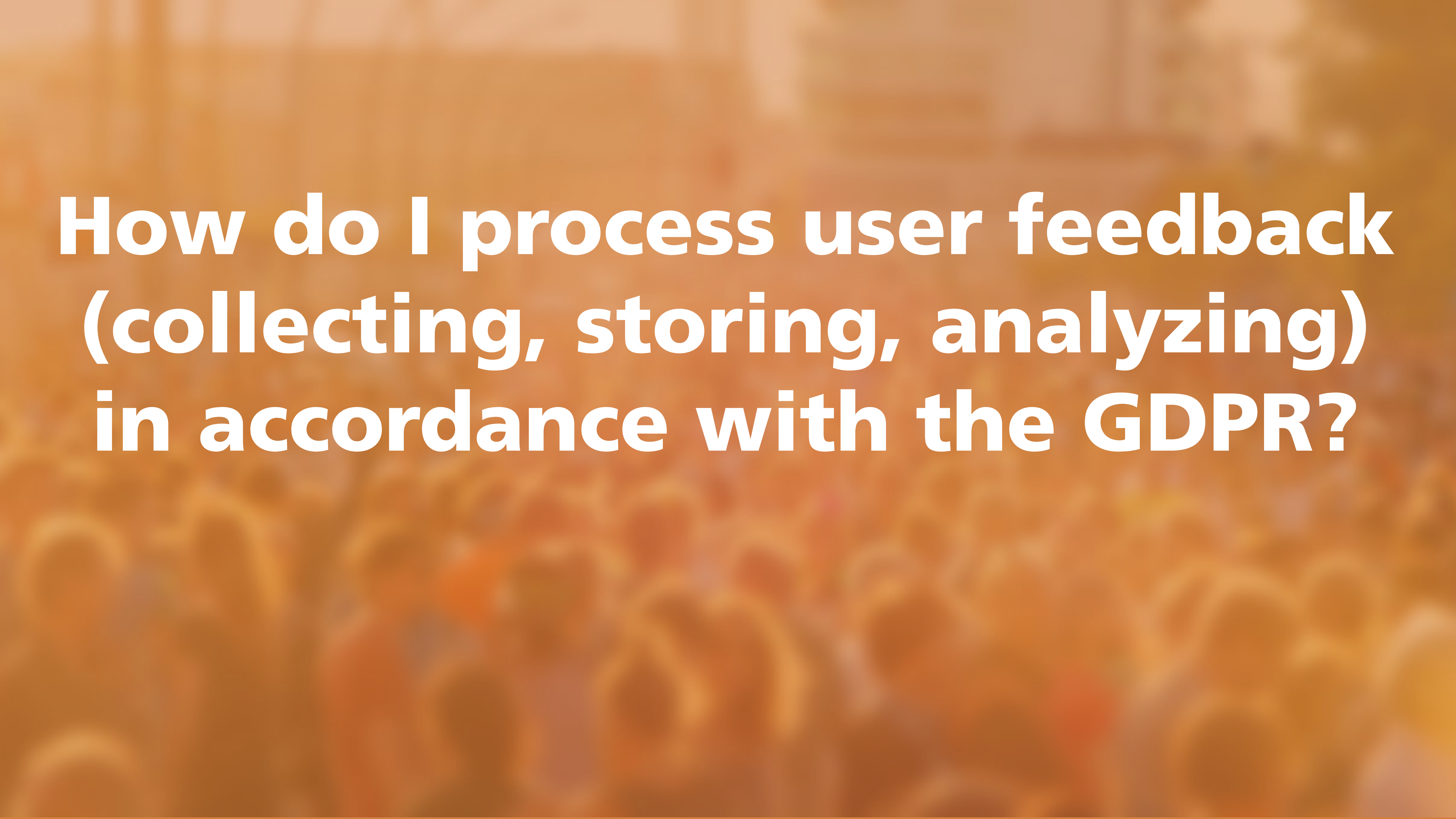
- **Home Address & Location**
  - *Search strategy:* typical words found in street names, and words such as “address”, “GPS”
  - *Result:* none found
    - Only matched unrelated aspects such as “for the road”, “road trips”, The only mentions of dates were recent, e.g., a date a review was updated; “street fighters” or “to address”

Findings 2/3

- **Name** (considered as first and last name, or as initial(s) and last name)
  - *Search strategy:* Manual inspection of the username variable
  - *Result:* Many matches, but always limited to the username
    - **About 4 in 10 user names on Amazon**
    - **About 1 in 10 user names on Apple App Store and Google Play**
    - Some even mention middle names and suffixes such as “Jr.”
- Some usernames are already anonymous
  - 4,477 “A Google User”; 292 “Amazon Customer”, 94 “Kindle Customer”, and 24 “Unknown”
- Some people may be using an alias that is different from their personal name
  - We should always assume that these names are personal data

Findings 3/3

- **Email address**
  - *Search strategy:* “@” signs
  - *Result:* We found email addresses and Twitter names
    - **Email addresses about 1 in 5,000 usernames**
    - **Email addresses about 1 in 300 user review texts**
    - Three occurrences of email addresses in review titles
      - Apparently posted by a spammer → violated rights of other users
    - Occasional mentions of the support email address contacted



How do I process user feedback (collecting, storing, analyzing) in accordance with the GDPR?

“Privacy can be tackled by certain motivation mechanisms, including assurance by the organization policies, and data protection measures, including the right of the crowd to know how their individual input was judged and by whom.

Duty to Inform

We note here that **such measures can become a burden** on the organization to adopt CrowdRE” *

* J. A. Khan, L. Liu, L. Wen, R. Ali, “Crowd intelligence in requirements engineering: Current status and future directions,” in Requirements Engineering: Foundation for Software Quality, LNCS 11412, E. Knauss and M. Goedicke, Eds. Cham: Springer, 2019, pp. 245–261.

Lawfulness of Processing Personal Data for CrowdRE

Art. 6 GDPR

1. Processing shall be lawful only if and to the extent that at least one of the following applies:

   (a) the data subject has given consent to the processing of his or her personal data for one or more specific purposes;

   (b) processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract;

   (c) processing is necessary for compliance with a legal obligation to which the controller is subject;

   (d) processing is necessary in order to protect the vital interests of the data subject or of another natural person;

   (e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller;

   Note: this can also be an end-user

   (f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child.

Contract

Possible Primary Constituents of Permission:

✓ GDPR Allows CrowdRE Analysis

Legitimate Interest

No legal counsel!

Consequence of Indirectly Obtained User Feedback: Duty to Inform (1/2)

- Personal data is not obtained directly from the data subject, but indirectly from another source
  - Then the 'controller' — the organization processing the data — to make a justified effort to inform the data subject about the use of personal data (Art. 14 (1) and (2) GDPR)
    - Within one month (Art. 14 (3) GDPR)
- Imposes additional costs for:
    - Informing data subjects
    - Putting procedures and mechanisms in place for granting the data subject's rights
      - E.g., a concept for sustainably deleting a data subject's data upon request
- Not doing so can cause severe fines (Art. 83 GDPR)

No legal counsel!

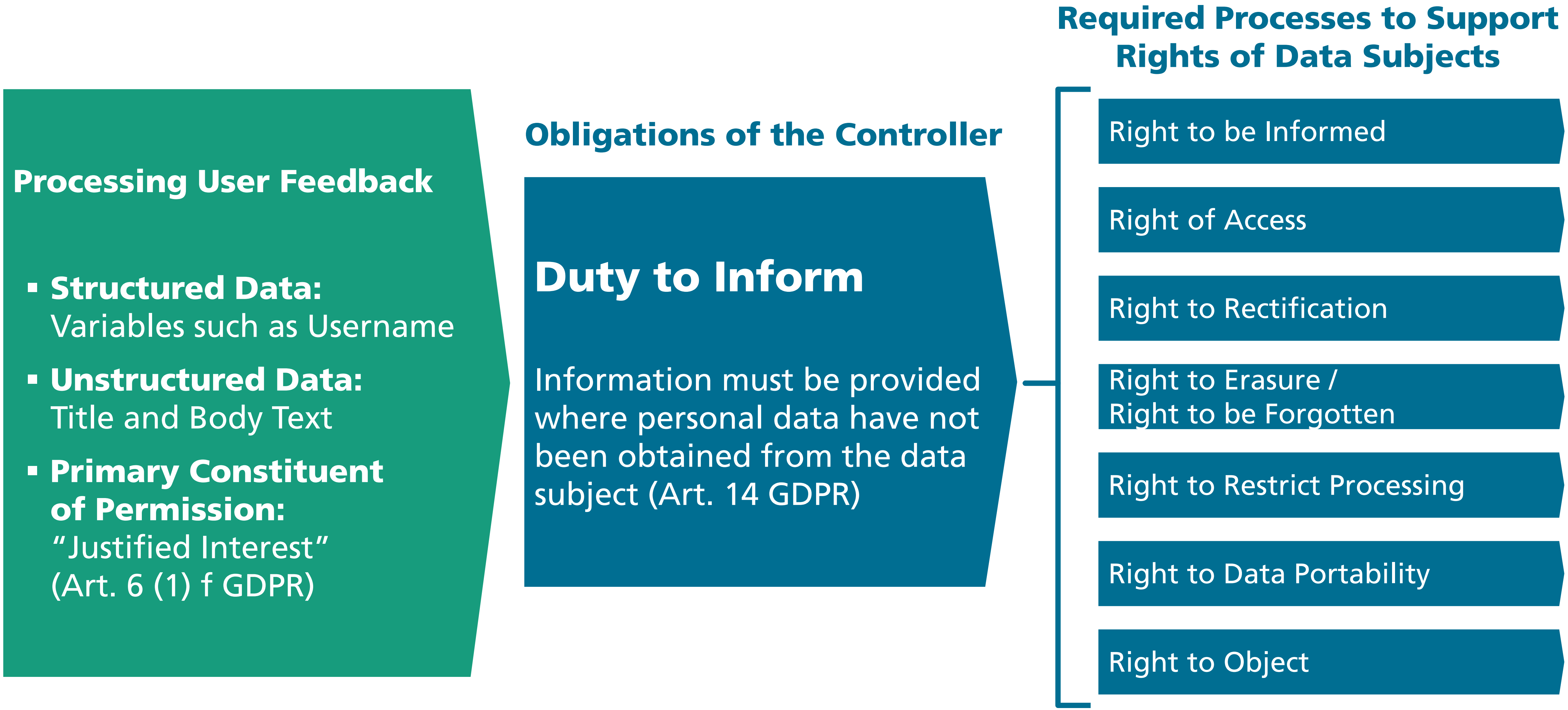
Consequence of Indirectly Obtained User Feedback: Duty to Inform (2/2)

- The message informing the user should among other things detail:
  - the organization providing the information,
  - the source of the data,
  - the purpose and legitimate interest of the processing,
  - the type of automated decision-making,
  - the recipients of the data (if any),
  - the duration for which the data will be stored,
  - the data subject's rights cf. Art. 15–21 GDPR).

- Risk of data subjects limiting or prohibiting the use of their data
  - Rendering that data useless
  - Requiring technical adaptations to manage the data in order to comply

No legal counsel!

Data Privacy in CrowdRE: Relevant GDPR Provisions for User Feedback



Processing User Feedback

- **Structured Data:** Variables such as Username
- **Unstructured Data:** Title and Body Text
- **Primary Constituent of Permission:** “Justified Interest” (Art. 6 (1) f GDPR)

Obligations of the Controller

Duty to Inform

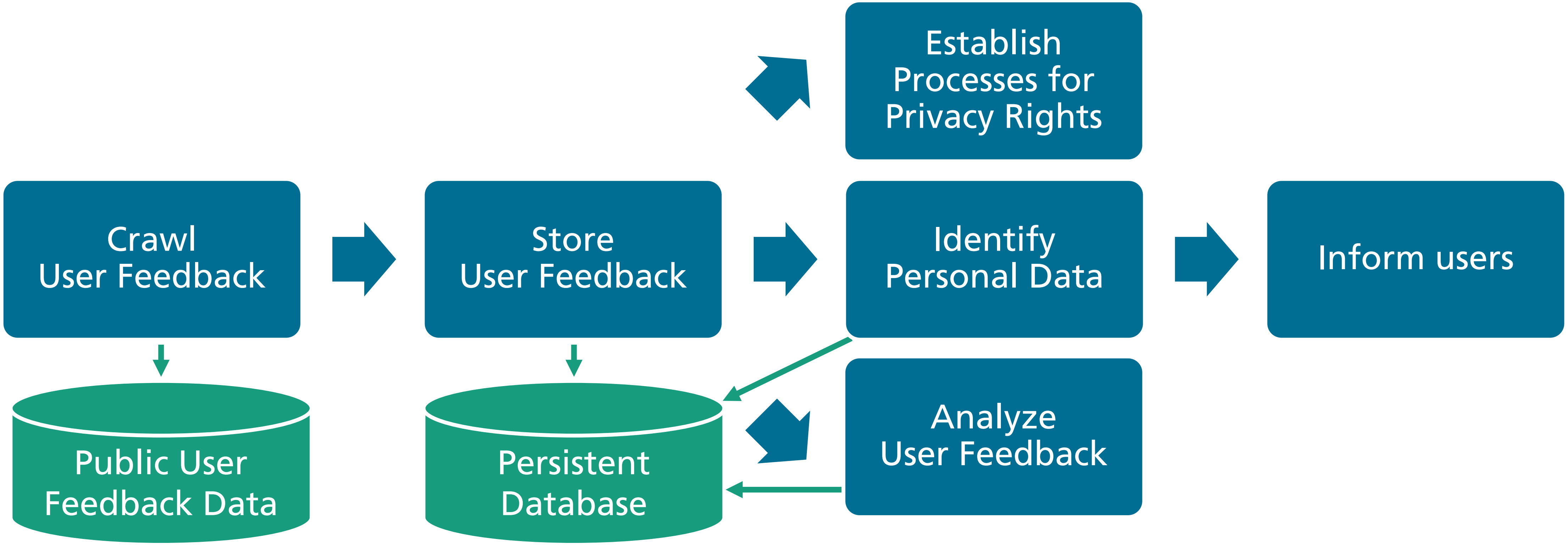
Information must be provided where personal data have not been obtained from the data subject (Art. 14 GDPR)

Required Processes to Support Rights of Data Subjects

- Right to be Informed
- Right of Access
- Right to Rectification
- Right to Erasure / Right to be Forgotten
- Right to Restrict Processing
- Right to Data Portability
- Right to Object

No legal counsel!

Process User Feedback for Commercial Purposes conform GDPR



No legal counsel!

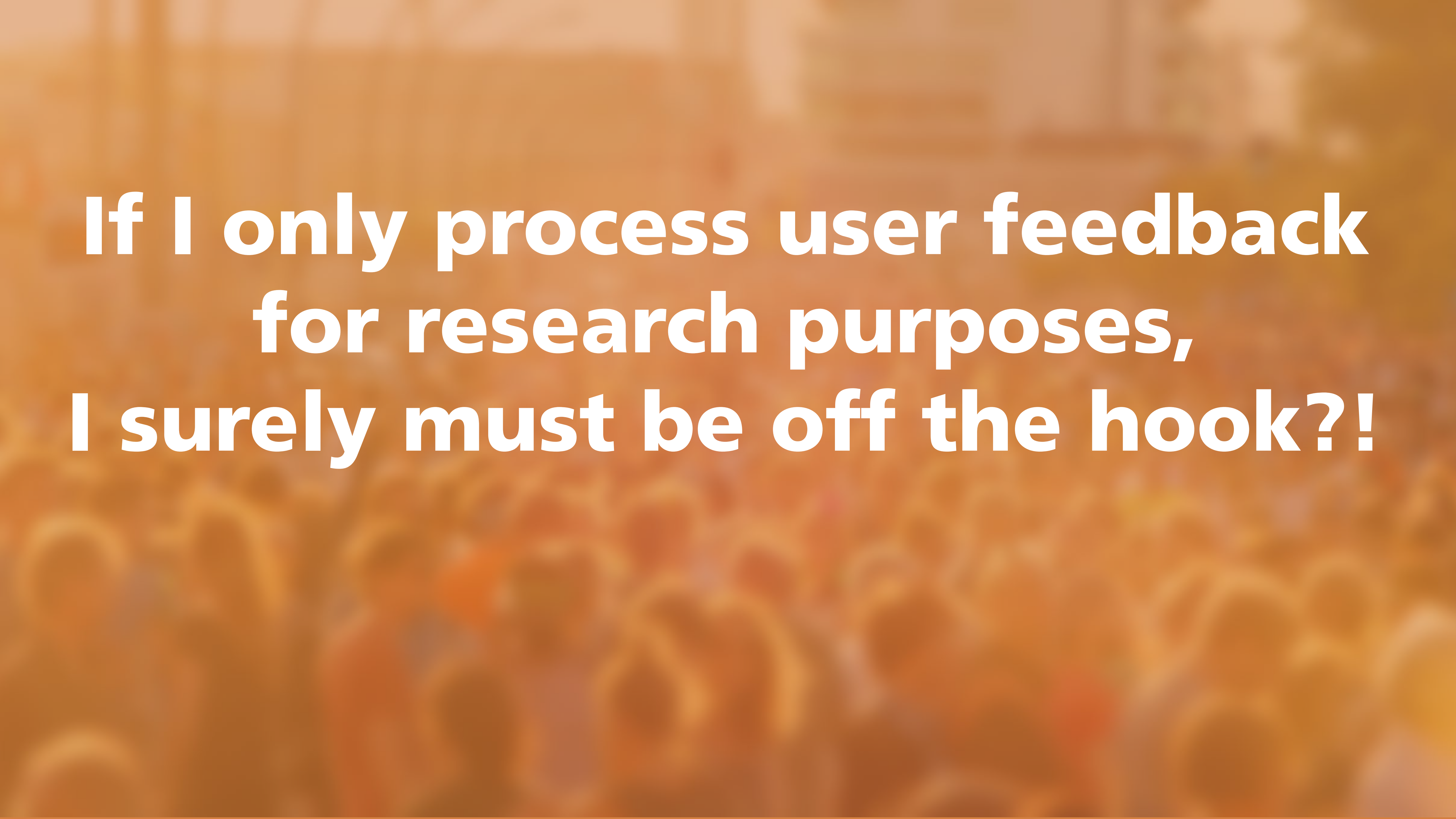
How Often? (An Example)

- We typically perform analyses over thousands of user reviews
  - E.g., competitor analysis including multiple apps

- Example: analysis of 15,000 user reviews
  - Contains the email addresses of over 50 persons
  - The party performing the analysis must inform all 50 persons
  - Additionally, the rights and freedoms of the data subjects must be ensured

- Justifiable disproportionate effort only if no valid contact information is found in the data
  - In all other cases, there is no exception: the data subject must be informed, and a processes covering Art. 15 through 21 GDPR to handle requests from data subjects needs to be in place

No legal counsel!



If I only process user feedback for research purposes, I surely must be off the hook?!

The bad news:
No. (Sorry!)

The good news:
Regulations are less strict.

Processing User Feedback for Research Purposes

Exemption from the Duty to Inform

Art. 14 GDPR

5. **The duty to inform** (…)

   (b) **shall not apply** where and insofar as (…) the provision of such information proves impossible or would involve a disproportionate effort, in particular for processing for archiving purposes in the public interest, **scientific** or historical **research purposes** or statistical purposes, subject to the conditions and safeguards referred to in Art. 89 (1) or insofar the obligation referred to is likely to render impossible or seriously impair the achievement of the objectives of that processing. In such cases the controller shall take **appropriate measures to protect the data subject's rights and freedoms and legitimate interests,** including making the information publicly available.

Special Measures

Challenge: Back-Searching

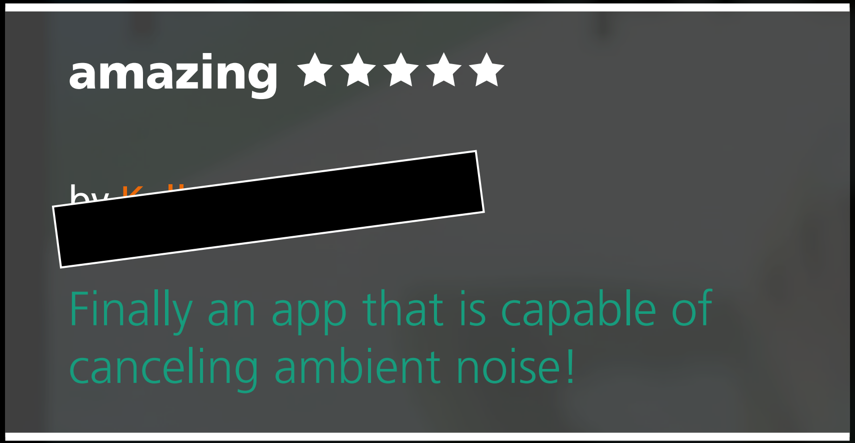
Art. 89 GDPR

1. (…) Those safeguards shall ensure that technical and organizational measures are in place in particular in order to ensure respect for the principle of **data minimization**. Those measures may include pseudonymization provided that those purposes can be fulfilled in that manner. Where those purposes can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects, those purposes shall be fulfilled in that manner.

Any way I can get around the GPDR?

Now we’re talking…

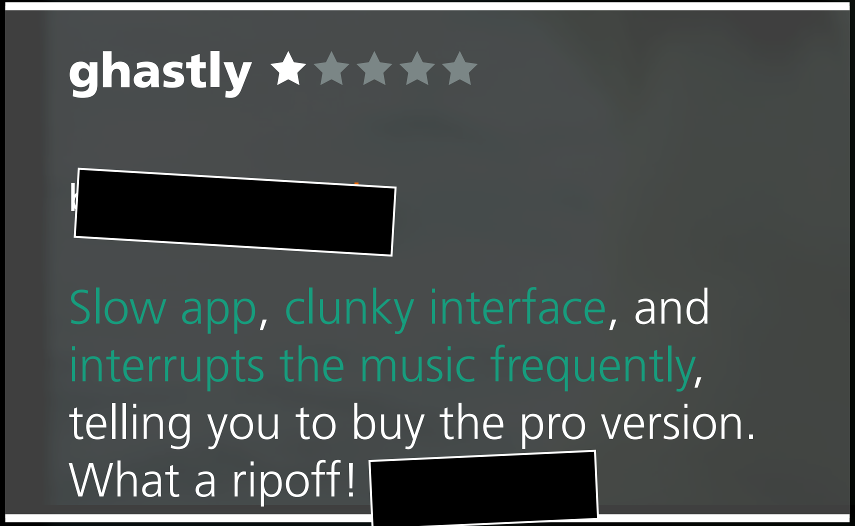
…anonymization.



amazing ★★★★★

by [redacted]

Finally an app that is capable of canceling ambient noise!



ghastly ★☆☆☆☆

[redacted]

Slow app, clunky interface, and interrupts the music frequently, telling you to buy the pro version. What a ripoff! [redacted]

Image source: FormatF Productions, used with permission

The GDPR does Not Apply to Anonymous Information

Recital 26 GDPR

The principles of data protection should therefore **not apply to anonymous information**, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. **This Regulation does not therefore concern the processing of such anonymous information.**

- Anonymous (anonymized) data is out of scope for the GDPR
  - GDPR regulations no longer apply to user feedback that is being stored and analyzed
    - No need to identify data subjects, inform them, or grant them their rights
  - Pseudonymized data that can no longer be attributed to a natural person even with the help of additional information can also be considered anonymized
- But: the user feedback must be anonymized or pseudonymized immediately and sustainably
  - Even storing the data for later anonymization would be a way of processing personal data and would thus be subject to the GDPR

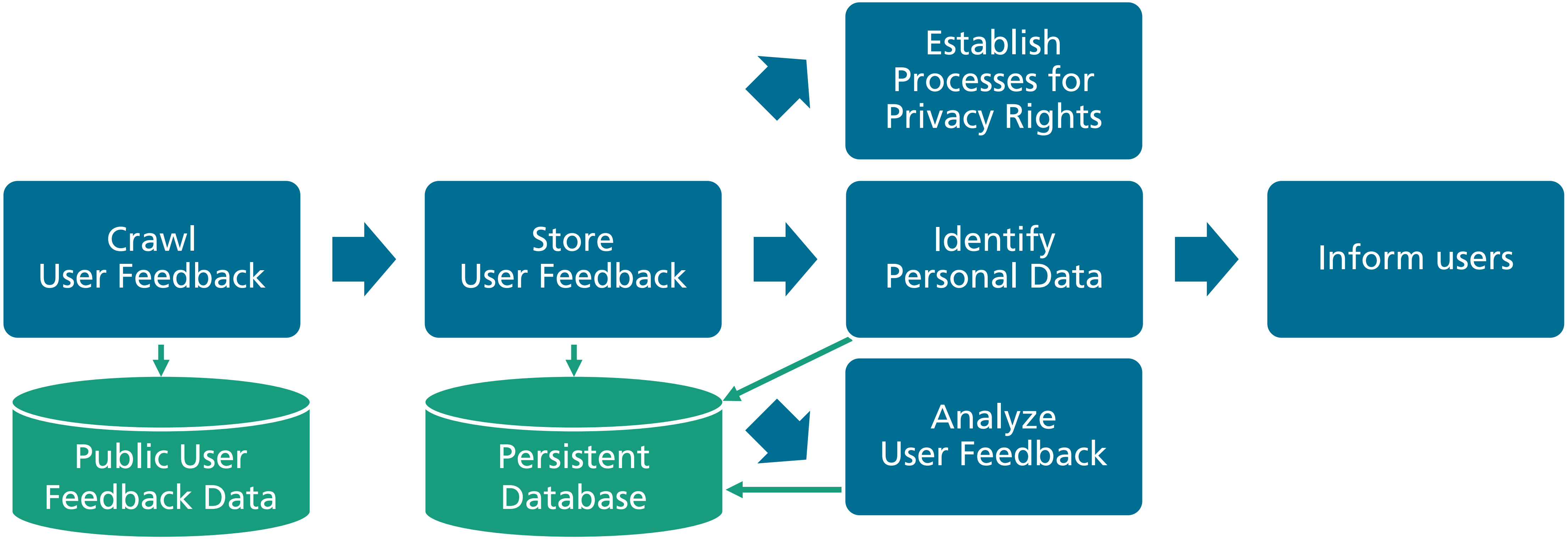
Proposed Solution Idea

Secured Environment

1. **Crawl** the user feedback
2. **Identify personal data** such as names and email addresses through, e.g.,
   - Artificial Intelligence (AI)
   - heuristics such as regular expressions or rule sets
3. **Anonymize** the identified personal data
4. **Persist** the anonymized data for CrowdRE analyses

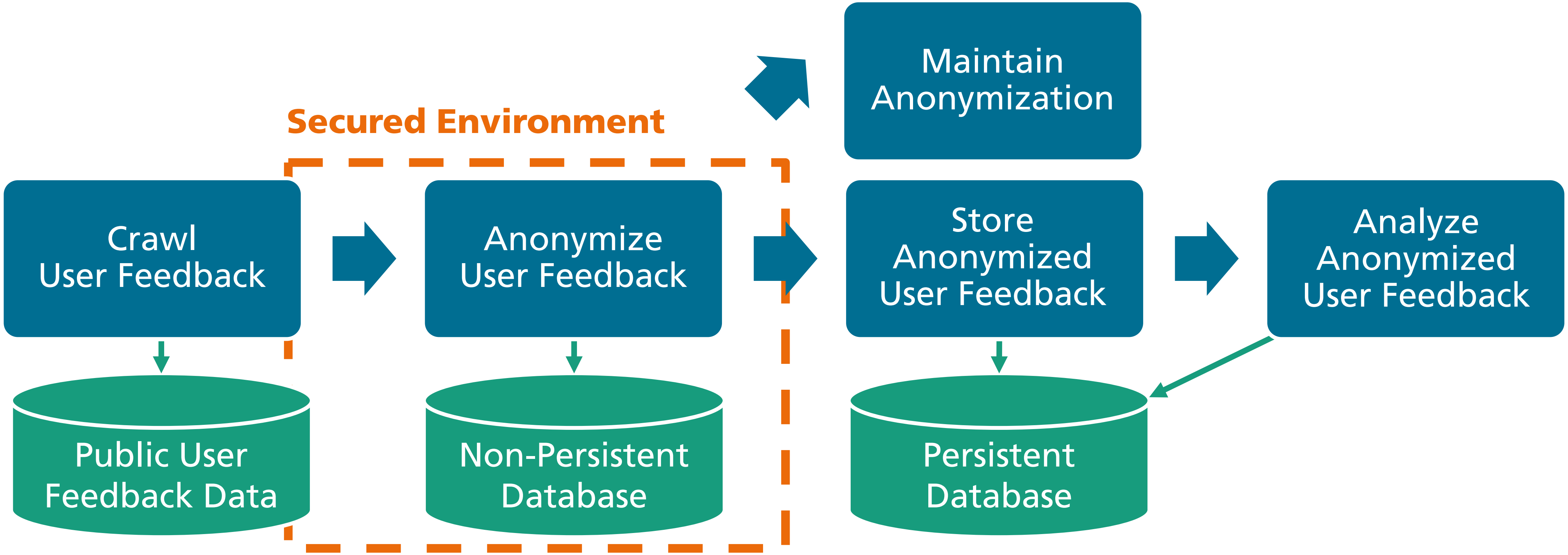
- Possible additional measures:
  - Put organizational measures in place, e.g., an organizational directive that prohibits back-searching the anonymized user reviews (e.g., using a search engine)
  - Set up a dedicated CrowdRE workstation with measures such as prohibiting users to access search engines and exporting raw data

Process User Feedback for Commercial Purposes conform GDPR



No legal counsel!

Process User Feedback for Commercial Purposes outside of GDPR through Anonymization



No legal counsel!

Anonymizing Email Addresses

- **A heuristic for identifying email addresses according to the pattern [user]@[domain].[extension] needs to be flexible and inclusive**
  - Account for use of spaces, domain name typos, placeholders for the "@" sign, etc.
- We cannot simply delete anything with an "@" sign:
  - "@" was used in 11% of the titles and 43% of the body texts, where it might be used:
    - to replace the word "at"
    - to censor use curse words, e.g., "Glitchy as !$&@/-"
- We cannot search simply for "email address" because we did not find co-occurrences of that word and the actual email address
- Twitter handles could be identified after email addresses are removed by identifying words and special characters structured like @[username]
  - Occasionally occur in combination with a mention of Twitter (e.g., "follow me on Twitter")

Anonymizing Personal Names

- First step: omitting the username variable
  - Or: pseudonymize user names (using one-way encryption)
    - One could detect anonymous names first and omit those from pseudonymizaiton
  - Unreliable measures to detect names: length, capitalization, blank spaces

- Means to identify personal names in the unstructured data of user feedback
  - Matching unstructured texts with a sub-string of the username
  - Searching for known names using a tool such as YAGO
    - They should account for foreign names and foreign writing styles

amazing ★★★★★

Finally an app that is capable of canceling ambient noise!

ghastly ★☆☆☆☆

Slow app, clunky interface, and interrupts the music frequently, telling you to buy the pro version. What a ripoff!

Image source: [FormatF Productions](), used with permission

Conclusions: CrowdRE, User Feedback and GDPR

- The GDPR allows user feedback analysis, but there is a duty to inform for commercial applications
- Technical and organizational measures can reduce/eliminate the impact of the GDPR on CrowdRE
- Next, we will (have to!) develop and fine-tune the heuristics for anonymizing user feedback

Technical Measures

Anonymization / pseudonymization of user feedback

Non-persistent (temporary) storage of user feedback prior to anonymization

Workstation inhibiting access to search engines and export of data

Organizational Measures

Organizational directives

Back-search prohibition

Data integrity policies

Regulations on accessing and using the user feedback database(s)

CrowdRE, User Feedback and GDPR

Towards Tackling GDPR Implications with Adequate Technical and Organizational Measures in an Effort-Minimal Way

Eduard C. Groen, Michael Ochs
Fraunhofer IESE

Eduard.Groen@iese.fraunhofer.de
Michael.Ochs@iese.fraunhofer.de

Jeju Island, South Korea
24 September 2019