Access rights to pseudonymous data

Experiences with Cloud providers

Sjoera Nas, IPEN, 9 December 2021
Agenda

• Experiences with many cloud providers
• Different categories of pseudonymous personal data
• Processor or joint controller
• Claimed exceptions on access rights
What cloud providers and services?

https://slmmicrosoftrijk.nl/downloads-dpias/
DPIA approach Privacy Company: legal and technical

Analysing framework agreement often many separate documents
Different kinds of Diagnostic Data

- **Telemetry data**: installed software applications and browsers collect and send data about the individual use of the services to the mothership.

- **Service generated server logs**: cloud providers register all end user activities in their own logs.

- **SIEM logs**: some personal data such as IP addresses in SIEM (security) logs.

- **Webserver access logs**: restricted and public access webpages.

- **Traffic to third parties**: cookies, support tickets, sometimes even Content Data (crash logs, spam/malware/CSAM scanning).
6 types of personal data

- Other natural persons
- 3rd parties
- Website data
- Content data
- Support data
- Contact & Account Data
- Diagnostic data (inc. security logs)
Inspection methods processing at cloud providers

- Laptop
- Smartphones: iOS, Android
- Online application

Visible → Invisible

Traffic interception → Interception and log inspection at provider → Data Subject Access Request at provider
Pseudonymously data are personal data

- Artikel 4(5) GDPR: 'pseudonymisation' means the processing of personal data in such a manner (...)
- Diagnostic data with unique identifiers are personal data
- Most cloud providers store diagnostic data for a long period of time in a single database. Hashing or other kinds of pseudonymisation do not help as companies can attach new information to the existing data with the same formula.
Example of Microsoft Windows 10 telemetry event

```json
{
  "ver": "3.0",
  "name": "Census.Enterprise",
  "time": "2019-01-23T07:12:19.6354125Z",
  "ty": "0MHzyskpyfwfg3p.0",
  "key": "0:ba893516ae74e80ae89c69d86d185e9ae3",
  "flags": 514,
  "ev": {}
}
```

**Standard header**

**Event specific data**

---

**Privacy Company**
All kinds of unique identifiers in telemetry events

```json
{
  "ver": "3.0",
  "name": "Census.Enterprise",
  "time": "2019-01-23T07:12:19.6354125Z",
  "cv": "B9E64b2b98fcafe4.0",
  "key": "s:0a8d8b518ae714e01ae89f96185e9ae3",
  "flags": 514,  
  "ext": 
    "pgName": "WINCORE",
    "flags": 91067332,
    "seq": 11
},

"BootId": 11,
  "name": "Windows",
  "ver": "10.0.17763.195.amd64fre.rs5_release.180914-1434",
  "app": 
    "id": "W:\f9d516ae714e01ae89f96185e9ae3",
    "ver": "2096/05/22:02:17:09!12DE0!devicecensus.exe",
    "flags": 0,  
  "device": 
    "localId": "s:0694CCC4-88C9-4A58-A0D0-7905553CE3F9",
    "deviceClass": "Windows.Desktop",
    "protocol": 
      "devMake": "VMware, Inc.",
      "devModel": "VMware Virtual Platform",
      "user": 
        "localId": "s:404E2543-61B-0X5-4722-9F664A31CD",
        "flags": 0,  
    "device": 
      "localId": "s:0694CCC4-88C9-4A58-A0D0-7905553CE3F9",
      "deviceClass": "Windows.Desktop",
      "protocol": 
        "devMake": "VMware, Inc.",
        "devModel": "VMware Virtual Platform",
        "user": 
          "localId": "s:404E2543-61B-0X5-4722-9F664A31CD",
          "flags": 0,  
  "data": 
    "IsCloudDomainJoined": 0,
    "IsMDMEnrolled": 0,
    "CDJType": 4294967295,
    "ServerFeatures": 
      "\0",
    "CommercialId": 
      "\0",
    "AzureIdentity": 
      "\0",
    "AzureOSSDPresent": false,
    "SAMLAssertion": "\0",
    "SystemCenterId": 
      "\0",
    "AADDeviceId": 
      "\0",
    "ContainerType": 4294967295,
    "EnrollmentType": "2494957295"}
```
<table>
<thead>
<tr>
<th>Time Stamp</th>
<th>Event</th>
<th>Event Location</th>
<th>Meeting ID</th>
<th>User ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>9/30/2020 12:19:40</td>
<td>Tap Security</td>
<td>In Meeting</td>
<td>focAptkITTSfHTiULJWSlw =</td>
<td>6n1pCAW4TT2qj5tmnGoKSg</td>
</tr>
<tr>
<td>9/30/2020 12:20:51</td>
<td>Recording</td>
<td>In Meeting</td>
<td>68460188777</td>
<td>6n1pCAW4TT2qj5tmnGoKSg</td>
</tr>
</tbody>
</table>

```json
{
  "client_os": "mac",
  "client_type": "Zoom Main Client",
  "client_version": "5.3.52877.0927",
  "event": "Recording",
  "event_loc": "In Meeting",
  "event_time": "9/30/2020 12:20:51",
  "meeting_id": "68460188777",
  "record": "toolbar-button",
  "sub_event": "Cancel",
  "user_id": "6n1pCAW4TT2qj5tmnGoKSg",
  "uuid": "3bdDCvtUdtgM7XuLZf79WlDlu7jTmLQ2YNzdDLg17A=",
}
```
Key problem: cloud provider as data controller
Joint controllership cloud providers with Enterprise customers

• Cloud providers often claim to act as data processor, but formal roles and contracts are not leading
• If a processor allows itself to determine processing purposes in its own interest, such as marketing or product innovation, it factually behaves as controller
• Most cloud providers omit to provide public documentation about the Diagnostic and Website Data
• The risk of factual, unintended joint controllership for the Diagnostic Data is that end users cannot exercise their data protection rights
Microsoft DSAR tools: audit logs, DDV and DSAR tool for diagnostic data

- **(Query on) audit logs**: always contain directly identifiable data, such as user ID, Organisation ID, IP address, subject header and addressees of mails, type of action + exact time

- **Diagnostic Data Viewer Tool**: shows very limited amount of telemetry events per app (less than 10%)

- **New DSAR tool for Diagnostic Data**: hodgepodge of undocumented files. Mostly with pseudonymised identifiers for the user and the tenant, such as "UserId": "b2814ab7-0cdb-4b7c-98fe-26deea388bb3"

- No access to **Required Service Data**

- No access to controller data (Windows, Edge)
Microsoft Diagnostic Data Viewer

### User Activity

<table>
<thead>
<tr>
<th>Date</th>
<th>Time</th>
<th>User Name</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>7-12-2021</td>
<td>12:19:39</td>
<td>userbi</td>
<td></td>
</tr>
<tr>
<td>7-12-2021</td>
<td>12:19:37</td>
<td>userbi</td>
<td></td>
</tr>
<tr>
<td><strong>scenario</strong></td>
<td><strong>12:17:16</strong></td>
<td><strong>userbi</strong></td>
<td><strong>Start</strong></td>
</tr>
<tr>
<td>7-12-2021</td>
<td>12:17:16</td>
<td><strong>scenario</strong></td>
<td></td>
</tr>
<tr>
<td>7-12-2021</td>
<td>12:17:04</td>
<td><strong>scenario</strong></td>
<td></td>
</tr>
</tbody>
</table>

### Network Activity

```json
{
    "complianceEnvironmentType": 0,
    "isDataCategorizationEnabled": true,
    "userpclevel": 3,
    "processMemory": 27963552,
    "freeMemory": 102535168,
    "clientType": "desktop",
    "batterylevel": 1,
    "pluggedin": true,
    "Window": {
        "Focus": "foreground",
        "Status": "systray",
        "Type": "main"
    },
    "windowIsVisible": true,
    "UserInfo": {
        "TimeZone": "+01:00",
        "Id": "234783b6-e703-4ba2-9772-60042325bab7",
        "TenantId": "439f01bd-ed3a-49cf-b758-9e0d41fe7526",
        "Ring": "general",
        "ETag": "\"gOKv1c6jJWOGa15XZ10qRPAwmmUN14t8Db7tZbx8Yg=\"",
        "Region": "emea",
        "LicenseType": "SmbNonVoice"
    }
}
```
This file contains 39 events recorded during the tests performed for this report. Example:

```json
{
  "time": "",
  "correlationId": "",
  "properties": {
    "Browser": "Firefox",
    "BrowserVersion": "91",
    "env_time": "2021-08-24 13:23:11.5473232",
    "IpAddress": "185.213.106.92",
    "ObjectId": "b2814ab7-0cdb-4b7c-98fe-26deea388bb3",
    "OfferId": "MS-AZR-0044P",
    "operationName": "SectionCompleted-PhoneVerification",
    "Platform": "Unknown",
    "Puid": "1003200138BoD6A3"
  }
}
```
AWS, Oracle, Microsoft Azure, Google Cloud VMs and databases

- Documentation and access for admins to detailed audit logs, but generally no individual take-out of end user data (system administrators)
- No access to data exchanged via market places (app stores)
- No access to cookies and data in webserver access log
- No access to all security logs that the provider creates behind the screens.
Claimed exceptions on access rights

- company confidential / non relevant internal information
- privacy of other data subjects
- technically too difficult
- not in a position to reliably identify the data subject
- referral to third parties
Dissection of refusal arguments

• **Company confidential**: If you can intercept personal data such as **telemetry events** in the network traffic, they cannot be company confidential (already made public). *Interception and access did reveal sensitive contents of telemetry data.* With regard to detailed **security logging**: the provider should provide access in the context of a DPIA to the customer's own data, and publish the results of an independent audit of the contents, retention periods and possible onward transfers to 3rd parties. *Access did reveal unknown 3rd parties (malware scanning)*

• **Cannot reliably identify**: must accept means such as identification in person, combination of timestamps with unique identifiers, bring along test device, provide full network capture, including identifiers
Dissection of refusal arguments

- **Too difficult to identify**: if the core business of a provider is to sell targeted ads based on these data, the categorical refusal to provide access is untenable. The provider must build the logs in such a way that access can be provided upon request without disclosing other individuals' personal data (Art 11 + Art 25 design)

- **Referral to third parties**: Customer/controllers must be able provide to access to traffic exchanged with third parties via their website. If they are joint controllers with for example Facebook for a Facebook Page, they must design a mechanism to provide access to these data
• Cloud providers must provide complete answers to DSAR, as controller and as processor (Art. 28(3) under e, f and h of the GDPR)
• If access to pseudonymous data is denied, the customer/controller cannot assess the scale and impact of the data processing
• Risk of factual joint controllership without division of responsibilities
• Lack of transparency about diagnostic data may be intentional or accidental: DSARs are necessary to assess accuracy of public documentation
• In view of the dynamic nature of telemetry data, end users should be able to get semi realtime access
Questions?

Contact me via www.linkedin.com/in/sjoera

www.privacycompany.eu
info@privacycompany.nl
070 – 820 96 90

Maanweg 174
Den Haag