Ministerie van Justitie en Veiligheid

DPIA Diagnostic data processing in Microsoft Office ProPlus

Reappraisal report with technical analysis

Version 1 – with summary in Dutch (met Nederlandse samenvatting)

Date	26 March 2019
Status	public

Colophon

DPIA by	**Ministry of Justice and Security** **Strategic Vendor Management Microsoft (SLM Rijk)** Turfmarkt 147 2511 DP Den Haag PO Box 20301 2500 EH The Hague www.rijksoverheid.nl/jenv
Contact	**Paul van den Berg** E p.j.van.den.berg@minvenj.nl> T 00 31 70 370 79 11
Project Name	**Follow-up DPIA report** with technical analysis of diagnostic data processing in Microsoft Office ProPlus
Authors	**Privacy Company** Sjoera Nas and Floor Terra, senior advisors www.privacycompany.eu

Change log

Version	Date	Summary of input
0.1	7 November 2018	First analysis of collected telemetry data with Fiddler, questions sent to Microsoft
0.2	14 November 2018	Microsoft agrees to investigate the specific telemetry events
0.3	17 December 2018	Between 12 and 16 December 2018 Privacy Company executes specific use scenarios of Office 365 CTR. SLM Rijk provides a Data Subject Access Request (DSAR) to Microsoft
0.4	15 January 2019	After a reminder about the DSAR Microsoft states it will not honour the request, because it has been hand-delivered by the SLM Rijk delegation, and not filed electronically.
0.5	12 March 2019	First completed draft with analysis of the collected telemetry events
0.6	13 March 2019	Added the used Word documents as appendix, input added from quick round of feedback SLM Rijk
0.7	14 March 2019	Input processed from SCC-I, with track changes
0.8	15 March 2019	Third draft, with editorial changes suggested by SLM Rijk, with track changes
0.9	15 March 2019	Third draft clean version for further distribution, personal data in the Annex anonymised
1	26 March 2019	Summary in Dutch added, URLs mentioned in footnotes revisited and recorded

Contents

Samenvatting

Dit rapport is geschreven in opdracht van de inkoopafdeling van het Rijk die verantwoordelijk is voor de aanschaf van Microsoft producten en diensten, SLM Rijk. Het is een herijking van de eerdere gegevensbeschermingseffectbeoordeling (Data Protection Impact Assessment, hierna: DPIA) van de verwerking van gegevens over het gebruik van Office 365 ProPlus (de software die op de apparaten van gebruikers wordt geïnstalleerd). Deze eerste DPIA is gepubliceerd op 7 november 2018. Dit tweede rapport moet als bijlage worden gelezen bij dit eerste rapport. Beide rapporten zijn geschreven door Privacy Company.

Office telemetrie software
Technisch gezien verzamelt Microsoft Corporation uit de Verenigde Staten diagnostische gegevens over het individuele gebruik van haar Office software via een ingebouwde Office telemetrie client. Microsoft heeft de Office software geprogrammeerd om telemetriegegevens te verzamelen op het apparaat, en de gegevens regelmatig naar Microsoft te sturen. De technische aanpak is vergelijkbaar met de ingebouwde telemetrie software in Windows 10, maar de Office telemetriestroom bevat andere gegevens. SLM Rijk adviseert overheidsorganisaties om de stroom Windows 10 telemetriegegevens te beperken door de (laagste) telemetrie instelling 'Beveiliging' te kiezen, maar deze telemetrie instelling heeft geen invloed op de gegevensstroom via Office.

Geen inzagesoftware in Office ProPlus
Dit herijkingsrapport bevat een analyse van de inhoud van de telemetriegegevens zoals die verzameld zijn door SSC-I, een IT-leverancier die deel uitmaakt van de Dienst Justitiële Inrichtingen. SSC-I heeft voor het ministerie van Justitie en Veiligheid een test lab ingericht. Het lab was niet in staat om de inhoud van de uitgaande datastroom te analyseren. Microsoft codeert het uitgaande verkeer naar haar servers, als essentiële beveiligingsmaatregel. Microsoft stelde nog geen software ter beschikking aan het lab om de gegevens te decoderen. Er is ook (nog) geen Diagnostic Data Viewer voor Office, zoals die wel beschikbaar is voor Windows 10 gebruikers sinds de lente van 2018.

Systeem-gegenereerde logbestanden
Microsoft's verzameling van diagnostische gegevens is niet beperkt tot de datastroom uit de ingebouwde telemetrie software in Office ProPlus. Microsoft verzamelt ook diagnostische gegevens op haar eigen cloud-servers, in systeem-gegenereerde logbestanden van gebeurtenissen. Microsoft legt bijvoorbeeld vast wanneer gebruikers documenten opslaan of openen in de cloud opslagdiensten SharePoint Online of OneDrive for Business (en andere clouddiensten van Microsoft die buiten het bereik van dit rapport vallen).

Na publicatie van de DPIA zijn er vier verschillende strategieën toegepast om meer informatie te krijgen over de inhoud van de verzamelde diagnostische gegevens.

In de eerste plaats zijn belangrijke delen van de verzamelde telemetriegegevens alsnog ontcijferd. Dit herijkingsrapport beschrijft de analyse van deze gegevens en geeft voorlopige conclusies over de risico's. In veel gevallen onderbouwt de analyse van de ontcijferde gegevens de beoordeling in de eerste DPIA dat Microsoft via de telemetriegegevens persoonsgegevens verwerkt.

Deze analyse geeft geen compleet beeld. Veel van de verzamelde gegevens waren gecodeerd op een andere, tweede manier, die niet zo makkelijk ontcijferd kon worden. En de telemetriegegevens zijn maar een onderdeel van de overkoepelende

categorie diagnostische gegevens. Daarom zijn er nog drie andere manieren verkend om een beter begrip te krijgen van de inhoud van de diagnostische gegevens die Microsoft verzamelt door het gebruik van Office ProPlus.

1. Microsoft is gevraagd om uitleg over de inhoud van specifieke telemetrieberichten and velden in die berichten, als de namen erop duidden dat het ging om een verwerking van persoonsgegevens
2. Er zijn nieuwe testscenario's uitgevoerd in Office ProPlus in een hybride omgeving (met toegang tot Sharepoint Online) en de betreffende onderzoeker heeft een inzageverzoek ingediend bij Microsoft als bedoeld in artikel 15 van de AVG.
3. Microsoft is gevraagd om, tijdens een bezoek van SLM Rijk aan het hoofdkantoor in Redmond, live inzage te geven in de diagnostische gegevens van een betrokkene die hiermee vrijwillig instemde.

Helaas heeft Microsoft niet meegewerkt aan deze verzoeken en geen uitleg gegeven over de inhoud van de verzamelde telemetriegegevens. Daarom is de reikwijdte van deze analyse beperkt.

Persoonsgegevens

Het oorspronkelijke Office ProPlus DPIA rapport bevat juridische, organisatorische en praktische argumenten waarom de verzamelde telemetriegegevens uit Office ProPlus persoonsgegevens zijn zoals bedoeld in art. 4(1) van de AVG. Dit tweede rapport met de technische analyse van sommige verzamelde telemetriegebeurtenissen bevat verder overtuigend bewijs dat de diagnostische gegevens persoonsgegevens zijn.

De analyse bevat duidelijke voorbeelden dat Microsoft de gegevens ook feitelijk gebruikt om gebruikers te identificeren. Microsoft gebruikt de diagnostische gegevens om individuele analyses te presenteren van werkgedrag aan werknemers en analyses te presenteren aan werkgevers, met de diensten MyAnalytics, Delve and WorkPlace Analytics.

Alle waargenomen telemetriegegevens kunnen vertrouwelijk zijn, of van gevoelige aard, afhankelijk van de omstandigheden. De tabel in dit rapport bevat voorbeelden van de gevoeligheid van de gegevens, maar de overheidsorganisaties die de Office ProPlus software gebruiken, moeten zelf identificeren in hun eigen DPIA's welke persoonsgegevens zij verwerken.

Inhoud en gevolgen van de waargenomen Office ProPlus telemetrie-gegevens

Dit rapport licht drie telemetrieberichten uit. Deze berichten bevatten unieke identificatoren die identificatie mogelijk maken van de individuele gebruiker, maar ze bevatten daarnaast ook informatie van gevoelige of vertrouwelijke aard.

Het eerste bericht toont aan dat Microsoft informatie verzamelt over de bestanden en padnamen die opgeslagen zijn in SharePoint Online en OneDrive for Business. Deze gegevens kunnen overheidsvertrouwelijke/gerubriceerde of zelfs staatsgeheime informatie of persoonsgegevens bevatten.

Werknemers hebben de gewoonte om hun eigen naam op te nemen in bestandsnamen, en organisaties kunnen werken met documentstructuren waarin de mapnamen vertrouwelijke informatie kunnen bevatten, zoals bijvoorbeeld beoordelingen of informatie over ziektes van de personen die in de bestanden in de map zijn opgenomen.

Het tweede bericht toont dat Word de inhoud van documenten scant om te beoordelen of het een CV is, en mogelijk informatie over de schrijver van het document combineert met informatie over de locatie en werkgever van die schrijver op Microsoft's eigen sociale netwerk LinkedIn. Microsoft vraagt niet om toestemming om bestanden voor dit specifieke doel te mogen scannen. Er is ook geen overduidelijke noodzaak voor Microsoft om dit soort informatie te verzamelen via een telemetriebericht om de (arbeids-)overeenkomst uit te kunnen voeren die overheidsorganisaties hebben met hun medewerkers.

Het derde uitgelichte bericht toont de URL van een online plaatje dat wordt ingevoegd in Word. Informatie over een bezochte webpagina is een persoonsgegeven van gevoelige aard, omdat een URL, los van de inhoud, informatie prijsgeeft over de interesses van gebruikers, en onderdeel is van de grondwettelijke beschermde inhoud van communicatie.

Dit rapport beschrijft nog andere telemetrieberichten waarmee Microsoft gedetailleerde informatie verzamelt over de inhoud van Word bestanden, zoals de aanwezigheid van een bibliografie, aantal tekens, alinea's, pagina's en plaatjes, evenals het aantal voet- en eindnoten. Microsoft verzamelt ook informatie over de nationaliteit van gebruikers (land, taal en taalinstelling op het systeem), over het individuele gedrag (clickstream) en activiteiten in Office (duur, gebruik in milliseconden en opgetelde gebruiksduur per activiteit in Office).

Andere gegevens die Microsoft verzamelt via de Office telemetrie zijn: het aantal mede-schrijvers uit de metadata van een bestand, het aantal of de bestandsnamen van ingevoegde bestanden, hoe lang een gebruiker heeft gekeken naar een gepersonaliseerde aanbeveling van Microsoft op het scherm en wat de persoonlijke interface voorkeuren zijn van een gebruiker in PowerPoint. Of de gebruiker het lint bijvoorbeeld aan heeft staan.

Samenvattend bieden alle geanalyseerde telemetrieberichten in dit herijkingsrapporat duiddelijke en feitelijke voorbeelden van de problemen en risico's die in de eerste Office ProPlus DPIA zijn beschreven met betrekking tot Microsoft's rol als verantwoordelijke voor de verwerking van gegevens via de 'vrijwillige' Connected Services, het gebrek aan doelbinding en grondslag,en de (dis-)proportionaliteit van de gegevensverwerking en privacy-onvriendelijke standaardinstellingen.

Conclusies

De analyse van de verzamelde telemetriegegevens in dit herijkingsrapport geeft duidelijke voorbeelden van de risico's die in de eerste DPIA zijn beschreven, en onderstreept de noodzaak voor overheidsorganisaties om maatregelen te treffen om de dataprotectie risico's voor betrokkenen te verlagen.

SLM Rijk heeft een handleiding opgesteld waarmee beheerders algemene maatregelen kunnen treffen om de risico's van de verwerking van diagnostische gegevens te beperken. Het gaat om een combinatie van de *zero exhaust* settings en het uitschakelen van functionaliteiten die doorgaans door Microsoft worden aangezet, zoals de 'vrijwillige' Connected Services.

Dit rapport adviseert beheerders om drie aanvullende maatregelen te nemen.

1. Als het niet mogelijk is om alle ‘vrijwillige’ Connected Services uit te zetten, om zoveel mogelijk losse diensten uit te zetten, in ieder geval de LinkedIn Resume Assistant
2. Verbiedt het gebruik van MyAnalytics
3. Maak geen gebruik van Workplace Analytics

Als gevolg van de eerste Office ProPlus DPIA heeft Microsoft zich eraan verbonden om verbeteringen door te voeren in de nieuwe versies van Office ProPlus die volgens schema in april 2019 gelanceerd zouden moeten worden. De verbeteringen hebben betrekking op transparantie en keuzemogelijkheden, in overeenstemming met de AVG-vereisten. SLM Rijk zal opdracht geven voor een nieuwe DPIA met analyse van de gegevensverwerking in de nieuwe april-versies van Office ProPlus 2019 en Office 365 Click To Run.

Summary

This report, commissioned by the Microsoft Strategic Vendor Management office (SLM Rijk) of the Dutch government, is reappraisal of the general data protection impact assessment (DPIA) on the processing of personal diagnostic data about the use of the Microsoft Office ProPlus software. The initial DPIA was published on 7 November 2018. This second-report has to be read as an annex to this first full DPIA report. Both reports have been written by the Dutch privacy consultancy firm Privacy Company.

Office telemetry client
Technically, Microsoft Corporation collects diagnostic data about the individual use of the Office software through the inbuilt Office telemetry client. Microsoft has programmed the Office software to collect telemetry data on the device, and to regularly send these data to Microsoft. The technical approach is comparable to the telemetry client in Windows 10, but it generates a different data stream. Government organisations are advised by SLM Rijk to limit the flow of telemetry data in Windows 10 Enterprise by choosing the *security* setting, but this Windows setting does not have any influence on the Office telemetry data stream.

No Diagnostic Data Viewer in Office ProPlus
This follow-up report contains an analysis of the contents of the telemetry data as collected by SSC-I, an IT-supplier that is part of the Dienst Justitiële Inrichtingen (the Custodial Institutions Agency). SSC-I has created a test lab for the Ministry of Justice and Security. The lab was unable to inspect the contents of the outgoing data stream. As an essential security measure, Microsoft encodes the outgoing traffic to its own servers. Microsoft did not provide tools to the lab to decode the outgoing data stream. Microsoft also does not (yet) provide a Data Viewer tool for Office, similar to the Data Viewer Tool that is available for users of Windows 10 since the spring of 2018.

System generated event logs
The collection of diagnostic data by Microsoft is not limited to the data stream sent through the in-built telemetry client in Office ProPlus. Microsoft also collects diagnostic data on its cloud-based servers, in system-generated event logs. For example, Microsoft records events when users store or access documents in SharePoint Online or OneDrive for Business (and other cloud-based services, out of scope of this DPIA).

After publication of the DPIA report four different fact finding strategies have been deployed to obtain more information about the contents of the collected diagnostic data.

First, relevant parts of the contents of the collected telemetry data were decoded. This follow-up report describes this analysis and draws provisional conclusions. In many cases the analysis of these decoded data proves the thesis in the first DPIA that Microsoft is processing personal data with the telemetry data.

However, many collected data fields were encoded in a second, different, way that was not easily decodable. And the telemetry data are a subset of the broader category of diagnostic data. Therefore, three other fact finding strategies were devised to gain a better understanding of the contents of the diagnostic data Microsoft collects through the use of Office ProPlus.

1. Ask Microsoft for clarification about the contents of specific events and fields in the collected telemetry data, with names that indicate the processing of personal data

2. Execute new scenario's in Office ProPlus, including access to SharePoint Online, and have an employee file a Data Subject Access request with Microsoft for the collected diagnostic data
3. Ask Microsoft to provide live access to diagnostic data from a consenting data subject during a visit from SLM Rijk to Redmond.

Unfortunately, Microsoft has not fully collaborated with these requests, and has not provided any input on the contents of the collected telemetry data. Therefore, the scope of this analysis is limited.

Personal data

The original Office ProPlus DPIA report provides legal, organisational and practical arguments why the collected telemetry data from Office ProPlus are personal data as defined in Art. 4(1) of the GDPR. This second report with the technical analysis of some of the collected telemetry events provides further conclusive evidence that the diagnostic data are personal data.

The analysis provides clear examples that Microsoft is factually identifying users in practice. Microsoft uses diagnostic data to present individual work behaviour analytics to employees and to employers, in the services MyAnalytics, Delve and WorkPlace Analytics.

All of the observed telemetry data can be confidential or sensitive, depending on the circumstances. Examples of the sensitivity are provided in the table in this report. It is up to the government organisations that use the Office ProPlus software to determine the specific personal data in their own data protection impact assessments.

Contents and impact analysis of the observed Office ProPlus telemetry data

Three events are highlighted in the report. They all contain unique identifiers that allow for identification of the individual user, but they also contain information of a sensitive or confidential nature.

The first highlighted event illustrates that Microsoft collects information about the names of files and pathnames stored in SharePoint Online and OneDrive for Business. These data can include government confidential/restricted/classified or state-secret data and/or personal data. Employees have a habit of including their own names in document titles, and organisations may work with document structures in which file paths may include confidential information or for example, qualifications or diseases of the natural persons mentioned in the documents in the folder.

The second highlighted event shows that Word scans the contents of documents to infer whether it is a resume, and possibly combines information about the author with information about the location and employer of that author on LinkedIn. Microsoft does not ask for consent to scan the document for this specific purpose, and there is no obvious necessity for Microsoft to collect this type of information via a telemetry message to perform the agreement between government organisations and their employees.

The third highlighted event reveals the URL of an online picture that is inserted in Word. Information about a visited web page is personal data of a sensitive nature, because a URL, regardless of the contents, reveals information about the interests of users, and is part of the (constitutionally protected) contents of communication

Additionally, this report describes other telemetry events through which Microsoft collects detailed information about the contents of Word documents, such as the presence of a bibliography, character, page, paragraph and picture count, the amount of end and footnotes. Microsoft also collects information about the nationality of users (country, language and locale of the system), about the individual user clickstream and detailed user activity in Office (duration, use in milliseconds and aggregated usage data per activity), the number of co-authors from the metadata pertaining to a document, the number or the file names of inserted files, how long a user has watched a personalised recommendation from Microsoft on screen, what the personal interface preferences are of a user in PowerPoint, whether the 'ribbon' is visible for example.

In sum, all the analysed telemetry events in this reappraisal report provide clear and factual examples of the problems and risks that have been described in the initial Office ProPlus DPIA regarding the qualification of Microsoft as a data controller for the 'voluntary' Connected Services, the lack of purpose limitation and legal ground, as well as the (dis)proportionality of the data processing and privacy unfriendly default settings.

Conclusions

The analysis of the collected telemetry events in this reappraisal report provides clear illustrations of the risks described in the first DPIA, and underlines the need for the measures organisations must take to lower the data protection risks for data subjects.

SLM Rijk provides a detailed manual for admins to generally lower the risks of the processing of diagnostic data, with a combination of the *zero exhaust settings* and the disabling of functionalities that are frequently switched On by default, such as the voluntary Connected Services. This report suggests that admins should take three additional measures.

1. If it is not possible to disable all voluntary Discretionary Services at once, specifically disable as many individual ones as possible, including the LinkedIn Resume Assistant
2. Prohibit the use of MyAnalytics
3. Do not use Workplace Analytics

As a result of the initial Office ProPlus DPIA, Microsoft has committed to make improvements in the new versions of Office ProPlus that are scheduled to be released in April 2019. The improvements relate to transparency and choice, in line with the requirements of the GDPR. SLM Rijk will commission a new DPIA report with an analysis of the data processing in the new April 2019 versions of Office ProPlus 2019 and Office 365 CTR.

Introduction

This report, commissioned by the Microsoft Strategic Vendor Management office (SLM Rijk) of the Dutch government, is a follow-up on the general data protection impact assessment (DPIA) on the processing of personal data about the use of the Microsoft Office ProPlus software. The initial DPIA report was presented on 6 November 2018 for a representative selection of government security, data protection and privacy officers. The report was published on 7 November 2018, with an update on the negotiations between the Dutch central government and Microsoft about the GDPR compliance.[1] Both reports have been written by the Dutch privacy consultancy firm Privacy Company.[2]

This follow-up report has to be read as an annex to this first full DPIA report. This follow-up assessment is intended to shed more light on the specific personal data that are processed via the telemetry client. A DPIA report does not provide qualifications about compliance with the law(s).

Umbrella DPIA

The Microsoft Strategic Vendor Management office (SLM Rijk) conducts the negotiations with Microsoft for the federal government, but the individual government organisations buy the licenses and determine the settings and scope of the processing of telemetry data by Microsoft Corporation in the USA. Therefore, like the initial DPIA, this follow-up DPIA is meant to help the different government organisations with the DPIAs they must conduct, but this document cannot replace the specific risk assessments the different government organisations must make. Only the organisations themselves can assess the specific data protection risks, based on their specific deployment, the level of confidentiality of their work and the types of personal data they process.

Limited scope

This follow-up report contains an analysis of the contents of the outgoing telemetry data as collected by SSC-I, an IT-supplier that is part of the Dienst Justitiële Inrichtingen (the Custodial Institutions Agency). SSC-I has created a test lab for the Ministry of Justice and Security (hereinafter: the test lab). In July and August 2018, the lab executed a number of specific scenario's in the four most widely used Office applications. The lab also used SharePoint Online and OneDrive for Business to store documents in Microsoft's cloud and used a few specific Connected Services, such as the Editor (online spelling checker), Insert Online Picture and Smart Lookup. The lab collected all outgoing data. The results were documented in a separate (government confidential) report.[3]

This update report will not repeat the extensive previous findings and assessments with regard to for example the qualification of Microsoft as a data processor or controller, or the legal grounds, or the guarantees with regard to the transfer of personal data to servers in the United States. This report also does not change the

[1] Rijksoverheid, DPIA op Microsoft Office ProPlus, 7 november 2018, URL: https://www.rijksoverheid.nl/documenten/rapporten/2018/11/07/data-protection-impact-assessment-op-microsoft-office

[2] https://www.privacycompany.eu/

[3] Departementaal Vertrouwelijk Rapport Microsoft Office - Verkeersstromen en Diagnostic Data SSC-I, 20 augustus 2018.

assessment of the high data protection risks or the mitigating measures the organisations that deploy the software may take to mitigate these risks.

This update report only provides limited insight in the telemetry events collected by Microsoft with the telemetry client in Office, due to two different sets of circumstances. First of all, the nature of the inspection with scripted scenarios only provides limited insight in the amount and contents of all the telemetry data Microsoft may collect about the use of Office ProPlus. Second, it was not possible to decode all the telemetry events and Microsoft has declined to provide explanations. The different circumstances are elaborated below.

Limitations of the technical set-up

The lab has only tested two versions and two specific local installations of two ProPlus Office products: the Office 2016 MSI and the Office 365 CTR.[4]

The scope of this report is limited to the processing of personal data via diagnostic data when using these two selected installed Office ProPlus versions running with the operating system Windows 10 Enterprise, with the Windows 10 telemetry setting set to the minimum level of 'security'.[5] This DPIA report does not provide an analysis of the data protection risks caused by the use of web-based version of Office 365 (also called Office Online). The data protection risks of the use of these cloud-based services will be assessed in another DPIA report.

Second, the lab performed limited activities, to ensure repeatability and reproducibility of the collected telemetry data. These scenarios were drafted to capture data from common use by government employees, but they are limited in time and scope. The scenarios involved the execution of a few scripted actions in each of the four most widely used Office tools (Word, Excel, PowerPoint and Outlook). These actions were activities such as opening and storing a document, sending an e-mail, including a picture in a PowerPoint, and misspelling a few words in Word. In the initial DPIA about Office ProPlus, Microsoft is quoted to explain that the company collects 23 to 25.000 different types of events via Office.[6]

Third, the process of collecting telemetry data is dynamic. Microsoft can add telemetry events on the fly, and collect other types of data, if the purposes comply with any of the 8 purposes described in the initial DPIA report. Therefore, this update report can only provide a snapshot of the telemetry data collected by Microsoft through the examined Office versions in the summer of 2018.

[4] Versions Microsoft Office Professional Plus 2016 MST 1806 (build 10228.20080 Click to run); and Microsoft Office 365 (Subscription Microsoft Office 365 ProPlus – Semi Annual Channel o version 1708 (Build 8431.2270 Click to run).

[5] SLM Rijk recommends to set the Windows 10 telemetry level to security since the summer of 2018.

[6] See footnote 1, the DPIA report on Microsoft Office ProPlus. Microsoft is quoted: "*Office telemetry contains between 23 and 25 thousand events, as opposed to 1.000-1.200 events for Windows 10. While Windows 10 telemetry is controlled by maybe 8 to 10 engineers, Office telemetry is in the hands of 20-30 engineering teams.*" Source: Meeting report 28 August 2018, answer to Q1.

Limited possibilities to understand the contents of diagnostic data

This follow-up report contains an analysis of the contents of the telemetry data sent from the devices to Microsoft by the technical lab. The technical lab was unable, when the tests were performed, to inspect the contents of the outgoing data stream. As an essential security measure, Microsoft encodes the outgoing traffic to its own servers.

In the summer of 2018, when the tests were conducted, Microsoft did not provide tools to the lab to decode the outgoing data stream. Because of the dynamic nature of the telemetry data collection, and the dynamic nature of the configuration of a tool to inspect the contents of the telemetry events, SLM Rijk did not follow-up on Microsoft's offer on 22 October 2018 to use a new test-version of Office 365 ProPlus CTR with a betaversion of a tool to inspect the contents of the data.

Microsoft does not (yet) provide a public available Data Viewer tool for Office, similar to the Data Viewer Tool that is available for users of (Home, Pro and Enterprise versions of) Windows 10 since the spring of 2018.

Four fact finding strategies

After the publication of the DPIA report four different fact finding strategies have been deployed to obtain more information abou the contents of the collected telemetry data.

1. Privacy Company was able to decode relevant parts of the contents of the collected telemetry data. This follow-up report describes this analysis and draws provisional conclusions. In many cases the analysis of these decoded data proves the thesis in the first DPIA that Microsoft is processing personal data with the telemetry data.

   However, many collected data fields were encoded in a second, different way, that was not easily decodable. Therefore, three other fact finding strategies were devised to gain a better understanding of the contents of the telemetry data.

2. SLM Rijk has asked Microsoft by mail of 7 November 2018 to explain the contents of three specific events with contents that indicated the processing of personal data, and provide an explanation of other telemetry events contained in the table in paragraph 4.4 of this report. This strategy was envisaged because Microsoft kindly offered the assistance of its engineers to help with the interpretation of some telemetry events during a meeting at the Ministry of Justice and Security on 31 October 2018. Based on the event names and snippets of content, some other events were included in a list of events and data that could potentially include remarkable personal data. Upon receipt of the selected telemetry events, Microsoft agreed to provide the requested analysis, but asked for help to map the telemetry data to the usage scenarios described in the (government confidential) report from the technical lab. The lab helped Microsoft with the mapping, but shortly before the SLM delegation travelled to Redmond to discuss the compliance issues raised by the DPIA report, Microsoft informed SLM Rijk that it was not able to dedicate engineering time to the answering of these questions.

3. An employee of Privacy Company executed new scripted scenarios between 12 and 16 December 2018 in Office ProPlus and wrote a Data Subject Access request as specified in article 15 of the GDPR. On 17 December 2018, SLM Rijk presented the signed letter with this request to the Microsoft delegation in Redmond, together with proof of identity of the employee. The request

contained a written explict authorisation from the employee for the SLM Rijk staff to obtain, on her behalf, a digital copy of the diagnostic personal data relating to her behaviour in Microsoft Office ProPlus in the 5 days preceding the request. A reminder was sent on 15 January, but Microsoft has since declined to provide the requested access. Microsoft explained that it would not respond, becausethe request was presented on paper, with a copy of the passport of the requesting data subject, and not electronically. Such an argument does not fly under the GDPR, but in the interest of further collaboration with Microsoft, the employee has decided not to file a complaint with the Dutch data protection authority.

4. As a last strategy to collect facts, the SLM delegation has asked Microsoft to provide live access to collected telemetry data during its visit from 17 to 21 December 2018 to the Microsoft headquarters in Redmond. Microsoft did not grant this access.

In sum, Microsoft has not fully collaborated with the fact finding scenario's and has not provided any (other) input on the contents of the collected telemetry data.

Dialogue with Microsoft

Currently, Microsoft provides no documentation, settings or publicly available data viewer tool for the Office telemetry data.

As a result of the initial DPIA, Microsoft has committed to make improvements in the new versions of Office ProPlus that are scheduled to be released in April 2019, with regard to transparency and choice, in line with the requirements of the GDPR.

As stated in the public update about the negotiations with Microsoft, on 26 October 2018 agreement was reached on an improvement plan in which Microsoft undertook to adapt its products for use by the Dutch government in compliance with the GDPR and other applicable legislation. Microsoft has committed to submitting these changes for verification in April 2019.[7]

In the third week of December 2018, a delegation from SLM Rijk has visited Microsoft headquarters in Redmond for detailed discussions of measures to lower the data protection risks for data subjects whose personal data are processed via the Office ProPlus software.

The Dutch minister of Justice and Security has explained in a letter to members of parliament that Microsoft has agreed to make a number of improvements in the new April versions of Office ProPlus, such as creating a switch to limit the diagnostic data streams to Microsoft, and full access to the remaining data streams to Microsoft. Meanwhile, Microsoft has provided temporary technical information for government organisations to limit the telemetry flow, via a *zero-exhaust script*. SLM Rijk since distributes a technical manual with relevant settings and codes at the request of government organisations.[8]

[7] Rijksoverheid, Update on negotiations between Dutch central government and Microsoft on GDPR compliance, 7 November 2018, URL: https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/rapporten/2018/11/07/data-protection-impact-assessment-op-microsoft-office/Update+on+negotiations+between+Dutch+central+government+and+Microsoft+on+GDPR+compliance+November+7.pdf

[8] Brief Minister van Justitie en Veiligheid, Reactie op berichtgeving in de media over dataverzameling en opslag door Microsoft, 20 december 2018, URL: https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/kamerstukken/2018/12/20/tk

Microsoft has confirmed to the USA based e-zine Politico that the company will increase transparency about the Office telemetry data flow."[9]

SLM Rijk will commission a new DPIA report with an analysis of the data processing in the new April 2019 versions of Office ProPlus 2019 and Office 365 CTR.

-reactie-op-berichtgeving-in-de-media-over-dataverzameling-en-opslag-door-microsoft/tk-reactie-op-berichtgeving-in-de-media-over-dataverzameling-en-opslag-door-microsoft.pdf (URL last visited and recorded on 26 March 2019).

[9] "*Microsoft will take additional steps to make it easier for customers to understand what data needs to go to Microsoft to run our services and why, and where data-sharing is optional.*" Quoted in: Politico, Daniel Lippman, 'Microsoft to update Office Pro Plus after Dutch ministry questions privacy', 8 February 2019, URL: https://www.politico.eu/article/microsoft-to-update-office-pro-plus-after-dutch-ministry-questions-privacy/ (URL last visited and recorded on 26 March 2019).

1. Processing of telemetry data

1.1 Office telemetry client

Technically, Microsoft Corporation collects data about the individual use of the Office ProPlus software through the inbuilt Office telemetry client. Microsoft has programmed the (installed versions of the) Office software to collect telemetry data on the device, and regularly send these to Microsoft. The technical approach is comparable to the telemetry client in Windows 10, but it generates a different data stream.

In practice most government employees use the Microsoft Office software on devices with the Windows 10 Enterprise operating system. The Windows 10 telemetry client regularly collects event data about the use of apps on the device, including about the use of the Office software. The Office ProPlus telemetry data stream sent to the Microsoft servers in the United States is separate from, and independent of, the telemetry data stream generated by Microsoft Windows 10.

However, since the telemetry data are stored in the Microsoft Cosmos database, and this involves a very large amount of data, there could be an additional or higher risk if the Windows 10 telemetry data were combined with the separate diagnostic data collected about the use of the Office software. This report assumes that all government organisations follow the recommendation from SLM Rijk to set the level of telemetry to minimum, to the *security level,* thus preventing Microsoft from capturing rich events about the use of the different Office applications.[10] If the Windows telemetry were set to enhanced or full, through the Windows telemetry, Microsoft could also collect rich information about the individual use of Office applications such as Word, Excel, Outlook or PowerPoint.

1.1.1 Included: Voluntary Connected Services

As documented in the initial Office ProPlus DPIA report, Microsoft considers itself to be a data controller for the use of discretionary Connected Services. Currently, Microsoft offers 31 Connected Services, of which 17 are discretionary. In the current contract between Microsoft and SLM Rijk, the collection of data through the discretionary Connected Services is excluded from the agreed privacy protections in the Enrolment framework. If the end-user decides to use these services, Microsoft considers that it is a data controller, and may process the resulting personal data for its own 12 purposes, as outlined in the General Privacy Statement. The lab has tested the following discretionary Connected Services: Editor (spelling and grammar), the Office Store (to download a plug-in for Excel, called 'Microsoft Dynamics add in'), Translator, Smart Lookup and Online Pictures

1.1.2 Included: SharePoint Online and OneDrive for Business

The tests include the collection of diagnostic data in a hybrid set-up, when a user of a locally installed Office ProPlus uses an Office 365 account to access or store documents in Microsoft's cloud-based services SharePoint Online or OneDrive for Business.

[10] SLM Rijk recommends all government organisations to set the Windows 10 telemetry level to 'Security', following a DPIA report written by the Dutch privacy consultancy firm Privacy Management Partners, delivered in June 2018 (report not published).

2. Methodology

In dialogue with Microsoft, the technical lab has ensured that the collection of telemetry data collected through the use of Office ProPlus was reproducible and repeatable, for example by limiting the amount of actions, and waiting 30 seconds between each action.

The technical lab has collected the outgoing telemetry data from the VM's with Windows 10 Enterprise with the following tools:

- Fiddler 5.09 (a tool that enables inspection of traffic sent through an SSL tunnel)[11]
- Network Monitor 3.410 (a tool that enables inspection of network traffic)[12]

Detailed set-up of the use of the Networkmonitor:

- Start the relevant VM with RDP;
- Log in with a local account
- Start the Networkmonitor > create new capture tab > start;
- Start the specific Microsoft Office application
- Log in with an Office365 account;
- Make New document;
- Modify the Trustcenter according to the demands of specific scenarios, under 'Options Trustcenter > Trust Center Settings > Privacy options';
- Close the Microsoft Office application;
- Start the Microsoft Office application;
- Execute the scripted scenario. Wait 30 seconds between each scripted action.
- Once the script has been fully executed, stop the Networkmonitor.

Detailed set-up of the use of Fiddler:

- Start the relevant VM with RDP;
- Log in with a local account
- Start Fiddler;
- Start the specific Microsoft Office application
- Execute the scripted scenario. Wait 30 seconds between each scripted action.
- Once the script has been fully executed, stop Fiddler.

The lab has stored the logfiles from both tools. The lab has compared the network endpoints with the (little) information made publicly available by Microsoft.

As explained in the introduction, the resulting telemetry data only present a fraction of the possible telemetry data that Microsoft may collect through the use of Office ProPlus. While Microsoft has explained that the company may collect between 23 and 25.000 types of telemetry events.

It has not been possible to decode all the collected telemetry events. Because Microsoft has not published documentation with the exact specifications of the data format, Privacy Company has made a best effort to extract content from the collected telemetry data. This has resulted in partial telemetry data, where certain chunks of non-textual data might be missing.

[11] Fiddler information, URL: https://Fiddler.en.uptodown.com/windows (URL last visited and recorded on 26 March 2019).

[12] Network Monitor information, URL: https://www.microsoft.com/en-us/download/details.aspx?id=4865 (URL last visited and recorded on 26 March 2019).

In the table below, a textual display is provided of data that could be deciphered. It is certain that blocks of information are missing from the excerpts below, because only the human readable data are included. Binary data have been excluded.

Experience from the Privacy Company researchers with Windows telemetry data suggests that there are no 'empty' fields. Thus, whenever a field is shown below, it is plausible that this field contains information, even if the excerpt below does not show garbled text.

Privacy Company has unpacked the data that were sent to the specific telemetry endpoint mobile.pipe.aria.microsoft.com. The data were organised in a (self-made) Django database.

3. Contents and impact analysis of the captured telemetry data

A typical telemetry event contains a unique number, a timestamp, several unique identifiers for the end-user and/or the acount and/or the tenant and/or the device, and different bits of information about the use of the different Office applications.

Information in the telemetry events that starts with 'data', contains information about the contents of a specific activity or type of information in a specific telemetry event. All fields that do not start with 'data' contain generic telemetry information that is common for all telemetry events.

This report highlights three specific events. It is followed by a table with other events that are deemed to include personal data, often of a sensitive nature.

As explained in the pararaph 1.2.2, SLM Rijk has asked Microsoft for help in understanding the contents of the collected telemetry events, and the purposes for the processing of these data. By e-mail of 7 November 2018, SLM Rijk has asked Microsoft for clarification about the different telemetry events captured with Fiddler, as shown in the first three highlighted events and in the table in paragraphs 4.1 through to 4.4 in this report. Additionally, SLM Rijk has tried other fact-finding strategies, including filing a data subject access request.

Microsoft has not provided any explanation or answer. In view of the lack of information from Microsoft, the analysis below of the contents and purposes for which Microsoft processes these data remains a bit speculative. Due to the nature of the decoding, some contents may have been misinterpreted.

The snippets quoted below are not reproductions of complete events; the beginning and ending of events may be missing.

3.1 Filepath and location

3.1.1 Event name: office_powerpoint_ppt_desktop_filesave[13]

```
Event.Source
EventInfo.InitId$4B755525-9B25-429B-8B67-4EDB88F83946
EventInfo.Name&office_powerpoint_ppt_desktop_filesave
EventInfo.SdkVersion'ACT-Windows Desktop-C++-no-1.7.278.1-no
EventInfo.Sequence
EventInfo.Source
EventInfo.Time
2018-08-02T10:31:23.623Z
PipelineInfo.AccountId e97859101370486fbcb67f5a023da3fc
Release.Audience
Production_DC
Release.AudienceGroup
Production
Release.Channel
Release.Fork
1802-Feb
Session.Id$4d748701-b9b3-413b-9d61-ff269b7cea17
User.TenantGroup
Commercial
UserInfo.TimeZone
+02:00
eventpriority
Data.FullSave
Session.MeasuresEnabled
Data.PPTSaveTime
Event.Time
App.InstallType
Data.FilePath
Data.FileSize
Data.MergeStarted
Data.SaveFromLocation
Data.SaveToLocation
Data.SaveType
Data.TimeInMs
Event.Flags
Event.SchemaVersion
Event.SequenceN
EventInfo.CRC32
Legacy.AppId
Legacy.OsEnv
zP.Data.CorrelationId
zP.Data.FileHash
zP.Data.FilePath
zP.Data.FileSize
zP.Data.FirstErrorTag
zP.Data.FullSave
zP.Data.LastTag
```

[13] As found in the telemetry data about the the scripted use of PowerPoint, both in Office 2016 and in Office 365.

zP.Data.MergeStarted zP.Data.PPTSaveTime **zP.Data.SaveFromLocation** **zP.Data.SaveToLocation** zP.Data.SaveType zP.Data.TimeInMs Event.SampleRate **Session.SamplingClientIdValue** **Client.Id**

3.1.2 *Impact analysis: file names and path names in SharePoint Online / OneDrive for Business*

This event is selected because it includes several unique identifiers that allow for identification of the individual user (**Client.Id, Session.SamplingClientIdValue, PipelineInfo.AccountId** and possibly, depending on the contents, also via **zP.Data.CorrelationId**).
The event reveals information about the location of data. It follows from the test scenario's that the PowerPoint presentation has been opened and stored on Microsofts cloud services SharePoint Online and OneDrive for Business. Via the events **Data.FilePath, Data.SaveFromLocation, zP.Data.SaveFromLocation, zP.Data.SaveToLocation** and **Data.SaveToLocation**, as well as the events **zP.Data.FileHash** and **zP.Data.FilePath** Microsoft collects information about the names of file and the pathnames stored in SharePoint Online and OneDrive for Business.

File and path names may reveal confidential/restricted/classified or state-secret information and/or contain directly identifying personal data. These personal data may have a sensitive nature. Employees habitually include their own names in document titles. But the file paths may reveal sensitive or special categories of data, for example if an organisation that works with health data, has a document structure in which file paths may include names of diseases or medical expertise.

3.2 Word and LinkedIn

3.2.1 Event name: Office.Word.LinkedIn.NLGAugmentorResumeClassificationResult[14]

```
DeviceInfo.OsName
Windows Desktop
DeviceInfo.OsVersion
10.0
Event.Id'712a7673-9c27-44e8-9fa1-76ae5c8116a1.45
Event.Name;Office.Word.LinkedIn.NLGAugmentorResumeClassificationResult
Event.Rule
110277.2
Event.Source
EventInfo.InitId$FC62D750-21B9-43F8-BB8C-721043BD840C
EventInfo.Name;office_word_linkedin_nlgaugmentorresumeclassificationresult
EventInfo.SdkVersion'ACT-Windows Desktop-C++-no-1.7.278.3-no
EventInfo.Sequence
EventInfo.Source
EventInfo.Time
2018-07-17T05:39:32.951Z
PipelineInfo.AccountId bffd26d9e49b4b7db4cb4df3425812de
Release.Audience
Production
Release.AudienceGroup
Production
Release.Channel
Release.Fork
1805-May
Session.Id$712a7673-9c27-44e8-9fa1-76ae5c8116a1
User.TenantGroup
E_Not_Init
UserInfo.TimeZone
+02:00
eventpriority
Data.AuthorMatch
Data.JobCityMatch
Data.JobStateMatch
Data.JobTitleMatch
Data.JobZipCodeMatch
#Data.ResumeAssistantTaskPaneEnabled
Session.MeasuresEnabled
Data.CollectionTime
Event.Time
App.InstallType
!Data.LinkedInResumeClassification
Data.MainPdod
Event.Flags
Event.SchemaVersion
Event.SequenceZ
J8b77539e8ac1441da9e32ab3a8950e3c-96ad9b3e-495b-48a8-a452-636bc0660f33-
6890
ClientI
act_default_sourcei 1.7.212.1
```

[14] as found in telemetry about Word, only in Office 2016.

service_id
$ABFD57BE-0E86-405D-82A1-2CB252659E5C
)$A4DA7C41-E85D-4914-BB8C-F545061ACF60q
custom
produce_label
ActionInfo.ProtectionSource Azure RMS
AppInfo.Id
WINWORD
AppInfo.Office.Version
2016
AppInfo.Workload
Office Add-in
DeviceInfo.Id$21a81d7e-7b82-4511-a16b-bbf3ab837116
DeviceInfo.Make
VMware, Inc.
DeviceInfo.Model VMware7,1
DeviceInfo.NetworkCost Unmetered
DeviceInfo.NetworkType
Unknown
DeviceInfo.OsBuild
10.0.17134
DeviceInfo.OsName
Windows Desktop
DeviceInfo.OsVersion
10.0
EventInfo.InitId$B37BABE3-B2F9-4931-BFC8-3D73FFDC7749
EventInfo.Name
produce_label
EventInfo.SdkVersion'ACT-Windows Desktop-C++-no-1.7.212.1-no
EventInfo.Sequence
EventInfo.Source
act_default_source
EventInfo.Time
2018-07-19T16:27:11.485Z
PipelineInfo.AccountId 8b77539e8ac1441da9e32ab3a8950e3c
ProductInfo.TelemetryAllowed
Usage Only
TenantInfo.Id$c0bcfcc2-7e05-47bf-bcf2-5a69899bfb56
UserInfo.TimeZone
+02:00
UserInfo.UserObjectId$6d0d81dc-51f2-4bb6-b221-0551a6c121bc
eventpriority
UserInfo.Id0
localadmin@tbtmsdsk013
UserInfo.Principal0
i(geert.vd.ploeg@tbt.vernieuwingsdomein.nl
EventInfo.CRC32

3.2.2 *Impact analysis: combination of location and job title data with LinkedIn*

This event is selected because it includes a user principal **(geert.vd.ploeg@tbt.vernieuwingsdomein.nl)** and a number of unique identifiers that allow for identification of the individual user (**DeviceInfo.Id$21a81d7e-7b82-4511-a16b-bbf3ab837116, UserInfo.Id0 localadmin@tbtmsdsk013,**

TenantInfo.Id$c0bcfcc2-7e05-47bf-bcf2-5a69899bfb56, UserInfo.UserObjectId$6d0d81dc-51f2-4bb6-b221-0551a6c121b, as well as two different values for **PipelineInfo.AccountId**.

The event shows a relation between data from the Word document and the Microsoft owned social network LinkedIn.

The event name **Data.ResumeAssistantTaskPaneEnabled** points to the use of the voluntary Connected Service *Resume Assistant*.

Microsoft provides the following description of this Connected Service:
"*When you open a resume document, the LinkedIn Resume Assistant task pane opens. If you want to tailor your resume to a particular role or company, you can choose to receive a set of LinkedIn-powered examples and suggestions. Once your resume is complete, you will be given an opportunity to post it to LinkedIn.*"[15]

LinkedIn wrote a blog post about its integration in Microsoft Word. LinkedIn wrote: "*today we're bringing the power of LinkedIn into Microsoft Word with the launch of Resume Assistant. With this integration, you'll get the inspiration and resources to craft a compelling resume directly within Microsoft Word, and you'll see relevant job opportunities on LinkedIn that are personalized for you*."[16]

It is unclear why Microsoft activated the Connected Service *Resume Assistant* during the test perfomed by the lab. Microsoft did not show information about, or ask consent for the use of, the Resume Assistant during the execution of the scripted scenario's about the use of LinkedIn or the specific Connected Service LinkedIn Resume Assistant. Microsoft explains that the Resume Assistant is only available if the user has an Office 365 license, and therefore, also has an Office 365 account.[17]

This telemetry event may be the result of one of the following two scenario's:

1. Opening of an existing Word document in Office 365 CTR stored in SharePoint Online. See the Word document with the title 'Usecase5_original'. This (test) document describes a job opening at the test lab at SSC-I.
2. Opening of an existing Word document in Office 356 CTR stored in OneDrive for Business. See the Word document with the title 'Usecase6_original'. This (test) document describes a job opening for the fictive personality of Sinterklaas at the test lab at SSC-I.

[15] See footnote 1, (first) Office ProPlus DPIA report, 7 November 2018, Annex 1, p. 84.

[16] LinkedIn Official Blog, Creating Your Resume Just Got a Whole Lot Easier with Microsoft and LinkedIn. Land Your Next Job Opportunity with Resume Assistant, 8 november 2017, URL: https://blog.linkedin.com/2017/november/8/Creating-your-resume-just-got-a-whole-lot-easier-with-Microsoft-and-LinkedIn (URL last visited and recorded on 26 March 2019). The blog post continues to explain the different combinations of data that LinkedIn can make based on analysis of (public) profiles and job openings.

[17] Microsoft explains: "*To use Office 365 ProPlus, a user must have an Office 365 account and have been assigned a license.*" Microsoft, About Office 365 ProPlus in the enterprise, URL: https://docs.microsoft.com/en-us/deployoffice/about-office-365-proplus-in-the-enterprise (URL last visited and recorded on 26 March 2019). And with regard to the Resume Assistant, Microsoft explains specifically: "***Resume Assistant** is only available in Word 2016 with an Office 365 subscription.*" URL: https://answers.microsoft.com/en-us/msoffice/forum/all/linked-in-resume-assistant-not-displayed-in-word/32e6f203-ac2e-4b3f-8119-9e605bb8ac31 (URL last visited and recorded on 26 March 2019).

The two word documents are included in the Appendix. The documents contain words such as 'vacature' (job opening), sollicitatie (job application) 'functie' en 'functieomschrijving' (function), 'kandidaat' (candidate) and 'salaris' / 'salarisniveau' (salary level).

One of the fields in this telemetry event is called 'Data.AuthorMatch'. Other fields seem to contain information about the location of the author (City, State and Zip Code match) and the job title of the author. It is not clear whether Microsoft compares or combines this information with information contained in the LinkedIn profile of that author.

From this unrequested activation of the Connected Service LinkedIn Resume Assistant three conclusions can be drawn.

First, the field names in this telemetry event seem to suggest that Microsoft matches information from the contents of a Word document to infer if the document is a resume. Microsoft does not ask for consent to scan the document for this specific purpose, and there is no obvious necessity for Microsoft to collect this type of information via a telemetry message to perform the agreement between government organisations and their employees. This event thus provides a clear and factual example of the problems and risks that have been described in the initial Office ProPlus DPIA regarding the lack of purpose limitation and legal ground, as well as the (dis)proportionality of the data processing.

Second, this event confirms that all Connected Services are switched On when a user chooses to use a particular Connected Service. In the two tested scenario's, only one specific Connected Service was used. But if a user once uses a Connected Service, such as the Editor (online spelling checker) in Word, all other Connected Services are switched on by default, in all Office applications, including Outlook and PowerPoint. Administrators can disable some of the individual voluntary Connected Services through group policies or with a reqistry key, but such a setting is not available for the Editor service. This provides a clear example of the problems with the privacy unfriendly design of the user interface in Microsoft Office, and the self-qualification of Microsoft as a data controller for the 'voluntary' Connected Services. Users are not able to give specific consent to specific types of data processing by the different Connected Services.

Third, the file names of the tested Word document were designed to be neutral. They do not provide any indication that the contents of the documents are job-or resume related. Microsoft explains that Word scans the contents of documents to detect whether a document is a resume. "*How does my resume get detected?*
Word scans for patterns in the documents you open, to determine if the document is likely to be a resume--similar to how grammar checking works. If you consent to use Resume assistant, then pattern-matched content from your resume is used to tailor the results in the Resume Assistant pane. For example, a job title and a location name allows for tailored job results. This is used only to enhance the Resume Assistant experience; Microsoft does not collect any personal information."[18]

[18] Microsoft, Write your best resume with help from LinkedIn and Resume Assistant, URL: https://support.office.com/en-us/article/write-your-best-resume-with-help-from-linkedin-and-resume-assistant-444ff6f0-ef74-4a9c-9091-ffd7a9d1917a?ui=en-US&rs=en-US&ad=US (URL last visited and recorded on 26 March 2019).

Microsoft confirms with this explanation that Office scans the contents of the Word documents to infer whether the document is a resume. The results of such a scan are evidently sent to Microsoft via this telemetry message, and possibly other telemetry messages that have not been observed.

The fact that Word has an undocumented feature to scan the contents of documents through an online Connected Service, involves high data protection risks for users. Even though Microsoft writes that users '*consent to use Resume Assistant*', during the lab test the testers were not asked for specific consent. They had only activated a specific Connected Service, but were not provided with any information about this purpose of the Resume Assistant service or processing. The data protection risks can be especially high for the datasubjects because Microsoft considers itself to be a data controller for the 'voluntary' Connected Services. Similarly, the possible combination of personal data from Word documents with data from LinkedIn profiles, also involves possible processing of personal data of a sensitive nature.

3.3 URLs

3.3.1 Event name: office_graphics_gvizinsertpicturetelemetry[19]

```
, 170007.3
Event.Source
EventInfo.InitId$3F579FF5-5778-450B-9CCE-99FACA72823B
EventInfo.Name*office_graphics_gvizinsertpicturetelemetry
EventInfo.SdkVersion'ACT-Windows Desktop-C++-no-1.7.278.1-no
EventInfo.Sequence
EventInfo.Source
EventInfo.Time
2018-07-19T15:16:42.442Z
PipelineInfo.AccountId cfcfdb91c68c4329bb8b7cb7babb3cf7
Release.Audience
Production_DC
Release.AudienceGroup
Production
Release.Channel
Release.Fork
1802-Feb
Session.Id$545b3435-e62b-4982-b29c-b72ad78762e1
User.TenantGroup
Commercial
UserInfo.TimeZone
+02:00
eventpriority
Session.MeasuresEnabled
Event.Time
App.InstallType
Data.PictureSize
Event.Flags
Event.SchemaVersion
Event.Sequence@
EventInfo.CRC32
Legacy.AppId
```

[19] As observed in Word telemetry in Office 2016 MSI and in Office 365 CTR.

Legacy.OsEnv **zP.Data.PictureFormatType** **zP.Data.PictureSize** **zP.Data.PictureSource**

3.3.2 *Impact analysis: collection of URL through voluntary Connected Service*

This event is selected because it includes at least one unique identifier, **PipelineInfo.AccountId cfcfdb91c68c4329bb8b7cb7babb3cf7**.

The event reveals information about an online picture that is inserted in Word. The insertion is done manually, not with the of the voluntary Connected Service Online Pictures. The event names **Data.PictureSize, zP.Data.PictureFormatType**, **zP.Data.PictureSize** reveal information about the type of picture, but the event **zP.Data.PictureSource** could reveal the URL of the picture. This reveals personal data of a sensitive nature, because a URL, regardless of the contents, reveals information about the interests of users, and is part of the (constitutionally protected) contents of communication.

3.4 Other remarkable and sensitive telemetry events

In the table below other remarkable events are marked in bold, with a short explanation of the possible contents of these fields.

Event name	**Remarkable content**	**Explanation**
GENERIC EVENTS FOR OFFICE		
office_system_**systemhealthmetadata**deviceconsolidated	zP.Data.ComputerSystemProductUuidHash* zP.Data.DeviceManufacturer zP.Data.DeviceModel zP.Data.DigitizerInfo zP.Data.IsLaptop zP.Data.IsTablet zP.Data.NumProcPhysCores zP.Data.NumProcShareSingleCache zP.Data.NumProcShareSingleCore zP.Data.PowerPlatformRole zP.Data.ProcSpeedMHz zP.Data.ProcTypeText zP.Data.ProcessorCount zP.Data.RamMB **zP.Data.SusClientId$** zP.Data.SysVolFreeSpaceMB zP.Data.SysVolSizeMB **zP.Data.WindowsSqmMachineId&** Event.SampleRate **Session.SamplingClientIdValue** **Client.Id** **Data.SusClientId**	This event collects information about the type of device contains different unique identifiers. This event proves that the telemetry data are personal data, because Microsoft or a third party (the tenant) are able to identify the individual user.

	Data.WindowsSqmMachineId **Session.ImpressionId** **User.TenantId** **)$AED5EB32-AB5F-40AD-905C-23BEAF3416F3q**	
Office_system_**systemhealthmetadata**applicationadditional	**zP.Data.CID** **zP.Data.CollectibleClassifications** **zP.Data.FirstRunTime** zP.Data.IsJoinedToDomain **zP.Data.IsLabMachine** **zP.Data.IsMsftInternal** **zP.Data.IsSubscription** **zP.Data.SqmUserId** **zP.Data.StudyId** zP.Data.WinUserActType Event.SampleRate Session.SamplingClientIdValue **Client.Id** **Data.SqmUserId** **Session.ImpressionId** **User.TenantId** **)$EA7F596B-A13E-4BE4-98FE-BD77B6C90587q**	This event contains different unique identifiers that disclose information about the nature of the user (UserId, ClientID, UserTenantID, Lab, Microsoft internal, Study ID) and use of the device (CollectibleClassifications, FirstRunTime). This event also proves that the telemetry data are personal data, because Microsoft or a third party (the tenant) are able to identify the individual user
office_system_**systemhealthmetadata**screencultureusersqmid	Data.IsJoinedToDomain **Data.IsLabMachine** **Data.IsMsftInternal** **Data.IsSubscription** Session.MeasuresEnabled **Data.CollectionTime** Event.Time App.InstallType Data.CollectibleClassifications **Data.CountryRegion** Data.HorizontalResolution **Data.KeyboardLanguage** **Data.OsLocale** **Data.OsUiLang** Data.ScreenDepth@ Data.ScreenDpi **Data.StudyId** **Data.SystemLocale** Data.TimeZoneBiasInMinutesw Data.VerticalResolution **Data.WinUserActType**	This event discloses information about the nationality, and as the name of the event suggests, *culture* of the individual user (with information about country, language and locale of the system). It is not clear what kind of information is collected in the field 'WinUserActType'.
Office_system_**system health**	**zP.Data.ActivityTag** **zP.Data.DurationInMilliseconds**	This event seems to collect information about the individual user

usage_nontcidclickstream	**zP.Data.InputType** **zP.Data.StartTime** **zP.Data.Success** **zP.Data.TelemetryId** **zP.Data.UserActionID**	clickstream, how long the user has provided input in what type of activity, with unique identifiers. Such clickstream data are considered **personal data of a sensitive nature**, as they can be used to analyse work patterns of employees.
office_fileio_csi_filerequestfinished	**Activity.AggMode** **Activity.Count** **Activity.Duration8** App.InstallType Data.Doc.AccessMode Data.Doc.EdpState Data.Doc.FileFormat Data.Doc.IOFlags **Data.Doc.Location** **Data.Doc.NumberCoAuthors**	This event discloses information about **detailed user activity and duration** in Office, as well as the **location of the requested file**, which may include file name and file path. **Both types of information are personal data of a sensitive nature**, as they may be used to analyse work patterns of employees. File and path names may reveal confidential/restricted/classified or state-secret information and/or contain directly identifying personal data. For example, if an employee uses his own name in a file name or file path. The event also collects **the number of coauthors** from the metadata pertaining to the document. Such metadata are **also data of a sensitive nature**.
office_media_insertmediadialogactivity	**Data.Dialog_Result** Data.Insert_Mode **Data.Inserted_Files**	With the field Inserted_Files Microsoft may collect the number or the **file names** of inserted files. The names of files are **data of a sensitive nature**, as they may be confidential/ restricted/classified information, and/or contain directly identifying personal data

office_ux_whatsnewtimeonscreen	**Data.WhatsNewOnScreenTimeSec**	This event may collect information that Microsoft has presented a personalised recommendation on screen to a specific user to explain what is new in Windows. With this event Microsoft records how long the user has watched this advertisement, before closing the pop-up. **This type of behavioural information, related to commercial purposes of Microsoft, is personal data of a sensitive nature**, as it can be used to profile the user and show him or her different recommendations.
act_stats	**Office_Telemetry_CustomProfile** Session.MeasuresEnabled App.InstallType EventInfo.CRC32 Legacy.AppId Legacy.OsEnv config_ecs_client_enabled config_inmemory_cache_size_bytes config_offline_storage_enabled !config_offline_storage_size_bytes config_version **9high_priority_log_to_successful_send_latency_millisec_min** **$high_priority_records_received_count** **)high_priority_records_received_size_bytes** **high_priority_records_sent_count** **2high_priority_records_sent_count_previous_sessions** +log_to_successful_send_latency_millisec_min	With this event, Microsoft collects information about **user activity** in Office ProPlus. This information could be used to map employee working patterns. This type of information can also be used to provide analytics to employers, as shown in Windows Analytics. It is not clear what the contents are of the field Office_Telemetry_CustomProfile. This could be a yes or no answer to the question whether the user has a custom profile, but it is not clear what a custom Telemetry profile could be.

	;normal_priority_log_to_successful_send_latency_millisec_min **&normal_priority_records_received_count.+normal_priority_records_received_size_bytes** **"normal_priority_records_sent_count04normal_priority_records_sent_count_previous_sessions** **0** **packages_count** **packages_succeeded_count** **record_size_bytes_max** **record_size_bytes_min** **records_received_count** **4** **records_received_size_bytes** **records_sent_count8$records_sent_count_previous_sessions8** **rm_bw_bytes_consumed_count** rtt_millisec_max rtt_millisec_min **session_start_timestamp** **Y** **session_startup_time_in_millisec** **stats_end_timestamp** **stats_send_frequency_secsx** **stats_start_timestamp**	
SPECIFIC POWERPOINT EVENTS		
office_powerpoint_docoperation_saveas	Data.DstDoc.Ext pptx **Data.DstDoc.Fqdn** sharepoint.com **Data.SrcDoc.Ext** pptx **Data.SrcDoc.StorageProviderId** **computer**	This event may contain information about the URL or file path (sourcedoc) of a stored PowerPoint presentation, including the name of the storage provider (in this case likely 'Computer'). The names of files and locations of files are **data of a sensitive nature**, as they may be confidential/ restricted/classified information, and/or

		contain directly identifying personal data
Name4Office.PowerPoint.PPT.Immersive.OARTActionsAggregate	**Data.Occurrence** **Data.Occurrence_0_50** **Data.Occurrence_1000** **Data.Occurrence_100_200** **Data.Occurrence_200_500** **Data.Occurrence_500_1000** **Data.Occurrence_50_100**	This event discloses information about (unknown OART) user actions in PowerPoint.
office_powerpoint_ppt_desktop_recordribbon	**Data.IsRecordPresenterViewLaunched** **Data.IsRecordingRibbonTabVisible** **Data.IsRibbonToolBarLoaded** Session.MeasuresEnabled **Event.Time** App.InstallType Data.RecordingRibbonTabGroupPolicyValue **Data.RecordingRibbonTabState**	With this event Microsoft collects information about the personal interface preferences of a user in PowerPoint, whether the 'ribbon' is visible for example. **This type of behavioural information, related to commercial purposes of Microsoft, is personal data of a sensitive nature**, as it can be used to profile the user and show him or her different recommendations.
office_powerpoint_ppt_hasusereditеddocument	**Client.Id** vn^F **Data.CorrelationId** **Session.ImpressionId** **User.TenantId**	Based on the name of this event (*has user edited document*), Microsoft collects personal data about user activity, with four unique identifiers.
office_powerpoint_docoperation_open (also found in different spelling with capitals, Office.PowerPoint.DocOperation.Open)	Data.AddDocumentToMruList. Data.CreateDocWindow Data.CreateDocumentToken **Data.DetachedDuration** Data.DetermineFileType Data.Doc.AccessMode Data.Doc.EdpState Data.Doc.FileFormat **Data.Doc.Location** Data.Doc.SessionId **Data.Doc.SizeInBytes** **Data.FileUrlLocation** Data.IncOpenDisabledReasons **Data.InitFileContents**	With this event Microsoft collects detailed information about the name and locations of files, and about the type and duration of user activity in PowerPoint, including information about the presence of coauthors. This information is **personal data of a sensitive nature**, **as it reveals detailed information about employee behaviour.** This type of information can be used

	Data.InitSecureReaderReasons Data.LoadDocument Data.OSRPolicy Data.OpenReason Data.SetDocCoAuthAutoSaveable **Data.StopwatchDuration** Data.SyncSlides **Data.UpdateCoauthoringState** Data.UpdateReadOnlyState Event.Flags Event.SchemaVersion Event.Sequence8 EventInfo.CRC32 Legacy.AppId Legacy.OsEnv Event.SampleRate **Session.SamplingClientIdValue** **Client.Id** vn^F **Data.Doc.UrlHash**	to provide detailed analytics to employers, as shown in Windows Analytics. **The locations and URLs of files are data of a sensitive nature**, as they may be confidential/ restricted/classified information, and/or contain directly identifying personal data
SPECIFIC WORD EVENTS		
office_word_experimentation_documentstatsoncloseandsuspend (also found in spelling with capitals, Office.Word.Experimentation.DocumentStatsOnCloseAndSuspend)	**Data.HasBibliography** **Data.HasHeader** **Data.IsImeUsed** **Data.IsPageCountInProgress** **Data.IsTouchUsed** Session.MeasuresEnabled Event.Time App.InstallType Data.BkmkRefCount **Data.CharacterCount** **Data.CharactersWithSpaceCount** **Data.ChartCount** **Data.CitationCount** **Data.DocumentLocation** **Data.EndnoteDocCount** **Data.FootnoteDocCount** **Data.IvyChartCount** **Data.LineCount** **Data.MainPdod** **Data.PageCount** **Data.PageNumberFieldCount** **Data.ParagraphCount** **Data.PicCount** **Data.RsidCount** **Data.TocCount** **Data.UserActionID**	This event collects detailed information about the contents of Word documents upon closing, such as the presence of a bibliography, character, page, paragraph and picture count, the amount of end and footnotes. This event also collects information about the time spent in milliseconds by the user, and the location of the document, which could include both file and pathname. **This information is personal data of a sensitive nature,** as it reveals detailed information about employee behaviour. This type of information can be used to provide detailed analytics to employers, as shown in Windows Analytics. The

	Data.UserInteractionTimeMsec **Data.WordCount**	locations and URLs of files are **data of a sensitive nature, as they may be confidential/ restricted/classified information**, and/or contain directly identifying personal data. The event also collects **information whether the user has used a touch screen**. In combination with the type of installed app. This could **disclose information about certain disabilities**.
office_word_fileopen_openloadfile	Data.BytesAsynchronous Data.BytesAsynchronousWithWork Data.BytesSynchronous Data.BytesUnknown Data.Doc.AccessMode Data.Doc.EdpState **Data.Doc.FileFormat** **Data.Doc.Location** Data.Doc.SessionId Data.Doc.SizeInBytes **Data.LinkStyles** **Data.MainPdod** Data.PartsUnknown Data.TemplateFormat **Data.UsesNormal** Event.Flags Event.SchemaVersion Event.Sequence6 EventInfo.CRC32 Legacy.AppId Legacy.OsEnv Session.Flags Event.SampleRate **Session.SamplingClientIdValue** **Client.Id** **Data.Doc.UrlHash**	This event collects the amount of data that are transferred when opening a Word file, the format of the file, the lcoation and the (hashed) URL. The locations and URLs of files are **data of a sensitive nature, as they may be confidential/ restricted/classified information**, and/or contain directly identifying personal data. It is not clear what the contents are of the fields MainPdod and UsesNormal.
office_word_fileopen_opencmdfilemrupriv	Data.BytesAsynchronous Data.BytesAsynchronousWithWork Data.BytesSynchronous Data.BytesUnknown **Data.DetachedDuration** Data.Doc.AccessMode	Similar to the previous event, this event also collects the location of a Word document, but additionally also collects information about the duration of activities,

	Data.Doc.EdpState Data.Doc.FileFormat **Data.Doc.Location** Data.Doc.SessionId Data.Doc.SizeInBytes **Data.MainPdod** Data.PartsUnknown **Data.StopwatchDuration**	with DetachedDuration and Stopwatchduration. This information is **personal data of a sensitive nature**, as it reveals information about employee behaviour.
Office.Word.Commanding.PasteFromExternalSPOSource	**Data.sourceFilePathHash** **Data.SrcDoc.UrlHash** **Data.FileHash@E3B0C44298FC1C149AFBF4C8996FB92427AE41E4649B934CA495991B7852B855**	Apparently, if a user pastes content in a Word document from an external source, Microsoft collects hashes of the source (**URL and file path) of the document. This is personal data of a sensitive nature**. Additionally, Microsoft collects **a hash of the contents of the file**. This can make the document identifiable if it concerns a public document. **With its own search engine, Microsoft** has already indexed a large part of the public internet and **is** thus **technically capable of calculating a hash**.

4. Personal data

The original Office ProPlus DPIA report provides legal, organisational and practical arguments why the collected telemetry data are personal data as defined in Art. 4(1) of the GDPR.[20]

1. Legal: the previous assessment by the Dutch data protection authority that the Windows 10 telemetry data are personal data.[21]

[20] Art. 4(1) of the GDPR: ***'personal data'*** *means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person*;

[21] Dutch DPA, report of findings Microsoft Windows 10, the processing of personal data via telemetry (in Dutch only), p. 101. URL: https://www.autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/01_onderzoek_microsoft_windows_10_okt_2017.pdf (URL last visited and recorded on 26

2. Organisational: the reasonable assumption of Microsofts' capability as a technology provider to be able to combine different telemetry events over time to identify a single user. Additionally, as part of its research, the Dutch DPA filed a data subject access request for its research accounts and established that it was factually possible for Microsoft to link the e-mail addresses to the user identifiers, and the user identifiers to device identifiers.[22]

3. Practical: the examples from the audits log about the use of the Office software in the test scenarios. According to Microsoft, the audit logs provide detailed information about product and service usage data contained in system-generated logs, such as the logs created by the use of Exchange Online, SharePoint Online en OneDrive for Business.[23]

The table in this follow-up report contains examples of collected telemetry events that provide more insight in the nature of the data that Microsoft collects via the in-built telemetry client in Office ProPlus. The events show that Microsoft collects several unique *identifiers* and information about the work patterns of employees, as well as confidential or sensitive personal data about the file names and locations when the Microsoft cloud storage services SharePoint Online en OneDrive for Business are used.

The table also shows that Microsoft collects additional information through its so called voluntary Connected Services, even if the service (such as the Resume Assistant) has not been requested or otherwise explictly permitted to operate. One of the telemetry events shows that Office has an in-built functionality to scan the contents of Word documents, and to send information to Microsoft about the results of this analysis via the telemetry events. Microsoft is technically capable of combining personal data included in a Word document with the personal data contained in the profiles of individuals on LinkedIn. This event provides another example of the ability of Microsoft to identify individual users through the telemetry events about their use of the Office software.
The table thus provides additional evidence that the Office ProPlus diagnostic data are personal data. Microsoft has explained that the company will include diagnostic data in the output of a Data Subject Request if they are personal data.[24]

Microsoft uses the diagnostic data collected through the use of Office ProPlus to provide three different kinds of analytic services. With MyAnalytics and Delve

March 2019). A summary in English is available at: https://autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/public_version_dutch_dpa_informal_translation_summary_of_investigation_report.pdf (URL last visited and recorded on 26 March 2019).

[22] Idem, p. 103.

[23] Microsoft, Office 365 Data Subject Requests for the GDPR, URL: https://docs.microsoft.com/en-us/microsoft-365/compliance/gdpr-dsr-office365?toc=/microsoft-365/enterprise/toc.json#part-2-responding-to-dsrs-with-respect-to-insights-generated-by-office-365 (URL last visited and recorded on 26 March 2019). Microsoft explains: "*Product and service usage data for some of Microsoft's most often-used services, such as Exchange Online, SharePoint Online, Skype for Business, Yammer and Office 365 Groups can also be retrieved by searching the Office 365 audit log in the Security & Compliance Center. For more information, see Use the Office 365 audit log search tool in DSR investigations in Appendix A.*"

[24] See footnote 1, first) Office ProPlus DPIA report, 7 November 2018, with reference to Meeting report 28 August 2018, answer to Q2.

Microsoft analyses individual work behaviour information and makes the insights accessible for each individual employee, but not for the administrator. Microsoft explains: "*MyAnalytics provides statistics to users to help them understand how they spend their time at work*"[25] and "*Delve uses intelligence to help employees discover relevant content and people across their organization. Users can only see documents they have access to*.[26]

In view of the stated purpose of these two analytic services, the collection of diagnostic data about the use of the Office ProPlus software has as content and as a purpose to identify the individual user. In 2007 the data protection authorities in the EU already explained that for data to relate to an individual, an important threshold for the qualification of personal data, there has to be a "content" element, OR a "purpose" element OR a "result" element.[27] Though the definition of personal data has been expanded in the GDPR to include for example location data and online identifiers, the analysis about the nature of personal data remains valid.

In the case of the different analytic services, the data meet at least the first two element. The DPAs write with regard to content: "*The "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.*
Information "relates" to a person when it is "about" that person, and this has to be assessed in the light of all circumstances surrounding the case. For example, the results of medical analysis clearly relate to the patient, or the information contained in a company's folder under the name of a certain client clearly relates to him."

With regard to purpose, the DPAs write: "*That "purpose" element can be considered to exist when the data are used or are likely to be used, taking into account all the circumstances surrounding the precise case, with the purpose to evaluate, treat in a certain way or influence the status or behaviour of an individual*."[28]

With Workplace Analytics, Microsoft provides organisations with insights about organizational productivity, collaboration patterns, and employee engagement. The processing is based on a user's email and calendar activities, plus additional data that an employer may choose to upload.

According to Microsoft, *Workplace Analytics contains aggregated, de-identified collaboration data of employees*.[29] It follows from a separate explanation about

[25] Microsoft, Office 365 Data Subject Requests for the GDPR, URL: https://docs.microsoft.com/en-us/microsoft-365/compliance/gdpr-dsr-office365?toc=/microsoft-365/enterprise/toc.json
[26] Microsoft, MyAnalytics vs. Workplace Analytics, Delve, and the Microsoft Graph, URL: https://docs.microsoft.com/en-us/workplace-analytics/myanalytics/overview/privacy-guide#myanalytics-vs-workplace-analytics-delve-and-the-microsoft-graph (URL last visited and recorded on 26 March 2019).
[27] Article 29 Working Party, WP 136, On the Concept of Personal Data, p. 10, URL: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf (URL last visited and recorded on 26 March 2019).
[28] Idem.
[29] Microsoft writes in its privacy guide voor MyAnalytics (see footnote 26): "*Although MyAnalytics is an individual productivity tool, Workplace Analytics enables organizations to view aggregated, **de-identified collaboration data of employees**.*".

system-generated logs that Microsoft defines de-identified data as pseudonymised data: "*Workplace Analytics also computes and stores* ***pseudonymized data*** *derived from Office 365 data to improve performance. If you would like to make* ***this pseudonymized data*** *available to a user and need assistance, contact Microsoft Support.*"[30]

Based on the definition in Art. 4(5) of the GDPR, pseudonymised data are personal data.[31]

Microsoft also alerts admins that their Workplace Analytics may contain personal data. "*Insights in Workplace Analytics reports created by you may or may not contain personal data of users that your organization licensed for Workplace Analytics, depending on the information that your organization used to supplement the Office 365 data. Your Workplace Analytics administrator will need to review those reports to determine if they contain a user's personal data. If a report does contain a user's personal data, then you will need to decide if you want to provide a copy of that report to the user. Workplace Analytics allows you to export the report.*"[32]

Employers are able through Workplace Analytics to analyse individual work patterns. Thus, the third element of 'result' in the definition of 'relating to' natural persons is met: "*Despite the absence of a "content" or "purpose" element, data can be considered to "relate" to an individual because their use is likely to have an impact on a certain person's rights and interests, taking into account all the circumstances surrounding the precise case.*"[33]

Though the data processing for these services is out of scope of the Office ProPlus DPIA and this update report, these analytic services provide clear illustrations how Microsoft uses the collected diagnostic data. Delve and Workplace Analytics are switched 'On' by default when using Microsoft Office 365.[34]

The likelihood of identifiability increases considerably with the ability to link different data events to an individual user. As explained in section 8 of the initial DPIA report, the processing of telemetry processing involves large amounts of data, with up to

[30] Microsoft, Additional steps to export system-generated log data, URL: https://docs.microsoft.com/nl-nl/microsoft-365/compliance/gdpr-system-generated-log-data (URL last visited and recorded on 26 March 2019).

[31] Art 4(5) GDPR: " **'pseudonymisation'** ***means the processing of personal data*** *in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person;*"

[32] See footnote 25, URL: https://docs.microsoft.com/en-us/microsoft-365/compliance/gdpr-dsr-office365?toc=/microsoft-365/enterprise/toc.json#part-2-responding-to-dsrs-with-respect-to-insights-generated-by-office-365

[33] Article 29 Working Party, WP 136, p. 11.

[34] In its privacy guide for MyAnalytics (see footnote 26), Microsoft explains that the Microsoft Graph cannot be switched off, but administrators can disable Delve and MyAnalytics separately. "*The Microsoft Graph* ***cannot be turned on or off globally*** *through the Office 365 Admin Center, but administrators can achieve this effect by blocking employees' ability to install third-party apps or by restricting developer access permissions.*" And: "*Administrators and individuals* ***can disable Delve content****-discovery functionality without impacting access to MyAnalytics, and vice-versa.*" And for admins: "*Use PowerShell to* ***opt employees out of MyAnalytics***".

25.000 different types of events. Microsoft has explained lots of engineering teams can add events to the data stream, while until recently there were no central rules governing the collection of the Office telemetry data.[35]

All the telemetry data (Office ProPlus and Windows 10) and the diagnostic data in system-generated server logs[36] are stored in the Cosmos database in the USA.[37] Both telemetry streams are dynamic. Following internal privacy approval, engineers may add new events to the stream. This *big data* processing makes it technically possible for Microsoft to create profiles of users and user groups based on the behavioural metadata collected over a period of time.

In sum, this follow-up report with technical analysis of some of the collected telemetry events provides further conclusive evidence that the diagnostic data are personal data.

5. Conclusions

The analysis of the collected telemetry events in this follow-up report does not change the initial analysis of risks and measures organisations must take to lower the data protection risks for data subjects.

Since January 2019, SLM Rijk provides a detailed manual for admins to generally lower the risks of the processing of diagnostic data, with a combination of the *zero exhaust settings* and the disabling of functionalities that are frequently switched On by default, such as the voluntary Connected Services.

General disabling of the voluntary Connected Services may however mean that some more advanced functions cannot be used. And central blocking is only possibly if the Office licenses are controlled locally via a KMS (Key Management Server)[38] or Active Directory based activation. Organisations may be required to change their Office licensing model in order to implement this measure.

1. If it is not possible to block of all voluntary Connected Services, organisations can lower the data protection risks for employees caused by the Connected Services by deploying the individual Group Policy settings Microsoft provides

[35] See footnote 1, first) Office ProPlus DPIA report, 7 November 2018, with reference to Meeting report 28 August 2018, answer to Q1. In its response to this DPIA Microsoft has explained that there are rules governing the collection of *new* telemetry events. See section 8 of the initial DPIA report.

[36] Microsoft confidential answers 1 October 2018 to the 10 follow-up questions, answer Q4f.

[37] "*Cosmos is the central audit record repository for all service teams and audit logs are uploaded to Cosmos from all servers in the Office 365 environment.*" Microsoft Compliance Manager Office 365, tab 'Microsoft Managed', Control ID: 6.9.3. Accessible (with Microsoft account log-in) via the Microsoft Servicetrust dashboard, the Compliance Manager, URL: https://servicetrust.microsoft.com/FrameworkDetailV2/b3d8589d-5987-45b7-8591-235c4a2f2ca2 (URL last visited and recorded on 26 March 2019).

[38] Microsoft, Activate volume licensed versions of Office by using KMS, URL: https://docs.microsoft.com/en-us/deployoffice/vlactivation/activate-office-by-using-kms (URL last visited and recorded on 26 March 2019).

for each of the different Connected Services.[39] To disable the specific Resume Assistant, admins can also use HKEY_CURRENT_USER\SOFTWARE\Microsoft\Office\16.0\Common\LinkedIn \OfficeLinkedIn=0.[40] However, Microsoft does not offer an option to specifically block the Editor service. Since this is probably the most frequently used Connected Service, an organisational measure such as the introduction of a request to users not to use the Editor, is probably not very effective. Therefore the general advice is to block all voluntary Connected Services, until Microsoft agrees to act as a data processor for these services, and agrees to make all the major changes to the data processing that are necessary to act in this (modest) role.

2. Use the setting 'excluded' for MyAnalytics. In that case, Microsoft promises: "*Office 365 data is not used for aggregated information shown to licensed users. User cannot opt-in through the Feature settings menu*."[41]

3. Do not use Workplace Analytics. There is no documented possibility to switch off the data processing by Microsoft for this purpose; admins can only decide not to use the functionality. In order to prevent accidental use of Workplace Analytics data, admins can set the aggregation level to include all employees.[42]

As a result of the initial Office ProPlus DPIA, Microsoft has committed to make improvements in the new versions of Office ProPlus that are scheduled to be released in April 2019. The improvements relate to transparency and choice, in line with the requirements of the GDPR.

From the viewpoint of SLM Rijk it is essential to reach agreeement as soon as possible with Microsoft about specific measures to mitigate the data protection risks. SLM Rijk also insists on public accountability, that all government organisations that deploy Microsoft Office ProPlus need to be able to verify the legal, technical and organisational changes to minimise the data protection risks for data subjects. It is essential that Microsoft makes available extensive public documentation about the contents and purposes of the telemetry events collected through Office ProPlus without any further delay.

SLM Rijk will examine the effectivity of the announced improvements. SLM Rijk intends to commission and publish a new DPIA report with an analysis of the data processing in the new April 2019 versions of Office ProPlus 2019 and Office 365 CTR.

[39] Microsoft, How admins can manage controller services in Office 365 ProPlus, 3 january 2019, URL: https://docs.microsoft.com/en-us/DeployOffice/manage-controller-services-office-365-proplus#resume-assistant (URL last visited and recorded on 26 March 2019).

[40] This registry setting has not been publicly documented by Microsoft, but has been reverse engineered through a differential comparison of registry key settings.

[41] Microsoft MyAnalytics setup for Office 365 Administrators, URL: https://docs.microsoft.com/en-us/workplace-analytics/myanalytics/setup/mya-setup-checklist#step-1-configure-user-settings (URL last visited and recorded on 26 March 2019).

[42] Microsoft privacy settings for Workplace Analytics, URL: https://docs.microsoft.com/en-us/workplace-analytics/use/settings#privacy-settings (URL last visited and recorded on 26 March 2019). Microsoft offers the following three options:
Specify the minimum group size
Specify whether to hide subject lines in meeting query results
Exclude words from subject lines

APPENDIX

WORD DOCUMENT USECASE 5
File name: Usecase5_original.doc

Existing Word document opened, the Connected Service used smart lookup, document stored in OneDrive for Business

Opengevallen functie SSC-I
Senior IT-Specialist Vernieuwing & Innovatie

Ministerie van Justitie en Veiligheid, Dienst Justitiële Inrichtingen (DJI)
SSC-I, locatie Zeist/Soesterberg (Schaal 11 BBRA), 36 uur per week

Functieomschrijving Het Shared Service Center voor ICT (SSC-I) – een grote ICT-leverancier binnen de Overheid – is op zoek naar een **senior IT-specialist met een duidelijke focus op vernieuwing en innovatie**. Wij zoeken iemand die er niet voor terugdeinst om af en toe écht out of de box te denken en met de allernieuwste technieken kan omgaan.

Je maakt deel uit van het Team Specials & Innovatie en werkt van daaruit zelfstandig of in het (project)team aan projecten en opdrachten. Je denkt als vanzelfsprekend mee over strategie en visie. De wensen van de klant zet jij om in een technische oplossing. De vakkennis die je als senior hebt, deel je binnen je team. De junior- en medior teamgenoten kunnen rekenen op jouw deskundige begeleiding. Innovatieve diensten hebben eigenlijk geen geheimen voor je. Jouw inbreng vormt een wezenlijke bijdrage aan doorlopende verbetering, niet alleen aan het tijdelijke beheer van de ontwikkelde diensten, maar ook aan vernieuwing in de bestaande dienstverlening.

Met ons team bouwen wij aan de beste oplossingen. Wij ontzorgen onze klanten, dat zit in ons afdelings-DNA.

Functie-eisen

Jij hebt HBO werk- en denkniveau, kennis van en ervaring met ITIL-processen en –werkwijzen en kunt terugkijken op minimaal zes jaar relevante werkervaring op het gebied van complexe technische infrastructuren.

Je hebt passie voor ICT-ontwikkelingen en ICT-beheer, je bent gedreven, initiatiefrijk, nieuwsgierig en enthousiast.

Van een senior verwachten we dat je je verantwoordelijkheid neemt en resultaat- en doelgericht te werk gaat. Je kunt zelfstandig en projectmatig samenwerken in een team, je bent klantgericht en hebt de beschikking over goede adviesvaardigheden, met name overtuigingskracht.

Wil je in deze functie kunnen floreren, dan moet je goed kunnen plannen en organiseren. Daarnaast ben je analytisch sterk, communicatief vaardig, een echte teamspeler en stressbestendig. Je vindt het leuk om jouw kennis te delen met teamgenoten en junioren en medioren op te leiden.

Een wezenlijk onderdeel van je taken wordt het vastleggen van ontwerpkeuzes in een technisch ontwerp (HLD en LLD) en het schrijven van werkinstructies voor collega’s. Dat betekent dat jouw kennis up to date moet zijn als we praten over gangbare standaarden op het gebied van apparatuur, netwerken, operating systems en middlewareproducten. Zonder die kennis kun je dat nooit vertalen naar de klantomgeving.

De functie vereist natuurlijk nogal wat technische kennis. We verwachten dan ook dat je niet schrikt van Red Hat, CentOS, Apache, PHP, MySQL en Opensource implementaties; Microsoft op MCSE- of MCITP-niveau; Citrix op CCA- of CCAA-

niveau; Scripttalen, bij voorkeur python en andere Linux script talen, VBScript en PowerShell; SQL Server, VMware op VCP-niveau; Kubernetes en Docker. En we gaan ervan uit dat je certificeringen up to date zijn én blijven;
En als je kennis hebt van OpenStack, GlusterFS en Lync / Skype for Business is dat ook nog meegenomen. We horen graag meer over je individuele specialisme(s).

Team Specials en Innovatie
Je komt te werken in het team Specials & Innovatie waar een deel van de integrale dienstverlening wordt gerealiseerd. Het team levert de dienstverlening voor onderdelen (producten en/of diensten) die nog niet volledig gestandaardiseerd zijn. Wij ontwikkelen en onderzoeken ook nieuwe vormen van dienstverlening door de integratie van de nieuwste technieken zoals bijvoorbeeld Docker,
Skype for Business, AI, VR etc. Denk hierbij ook aan specifieke klant-toepassingen zoals Elektronische Monitoring.

Verder ondersteunt de afdeling de innovatieve activiteiten van SSC-I door onze grote kennis van complexe en minder complexe infrastructuren. Daarnaast zijn wij verantwoordelijk voor beheer tijdens experimentele situaties, zoals bij een PoC/Pilot.

Wij zijn trots op onze flexibiliteit, onze korte time-to-market bij projecten en op ons hele werktraject: vanaf het standaardiseren en tijdelijk beheren van diensten tot aan de overdracht aan de beheerorganisatie.

Over SSC-I

SSC-I is dé ICT-dienstverlener van de Dienst Justitiële Inrichtingen van het ministerie van Justitie en Veiligheid. In totaal werken circa 450 vaste en plm. 125 externe medewerkers bij SSC-I. Het merendeel daarvan werkt in Gouda, een ander deel in de nevenvestigingen in Zeist/Soesterberg (jouw beoogde standplaats) en Veenhuizen. SSC-I in Gouda kent vier stafafdelingen en de directies Ontwikkeling & Vernieuwing en Beheer.

De regieorganisatie van DJI, Directie Informatisering (DI) is op dit moment de voornaamste opdrachtgever, waardoor DJI de grootste klant van SSC-I is.

SSC-I ontwikkelt, bouwt, implementeert en beheert ICT-werkplekken, applicaties en infrastructuur voor meer dan 20.000 gebruikers. Door de jarenlange ervaring binnen het veiligheidsdomein van de Rijksoverheid, zijn wij gespecialiseerd in veilige, betrouwbare en innovatieve ICT-oplossingen.
Buiten DJI bedienen wij diverse klanten binnen het veiligheidsdomein, zoals de Raad voor de Kinderbescherming (RvdK), de Dienst Terugkeer & Vertrek (DT&V), de Immigratie- en Naturalisatiedienst (IND) en Justis. Wij hebben de ambitie om de klantenkring binnen het ministerie van Justitie en Veiligheid verder uit te breiden.

Arbeidsvoorwaarden Salarisniveau: Schaal 11 BBRA. Min. € 2996 - Max. € 4605 (bruto per maand) Minimaal aantal uren per week: 36 Maximaal aantal uren per week: 36 Een consignatiedienst maakt onderdeel uit van de werkzaamheden. Contractduur: Vaste aanstelling (eventueel met een proeftijd)
Het genoemde salaris is gebaseerd op een volledige werkweek.

Overige arbeidsvoorwaarden Bovenop het salaris en vakantiegeld kun je rekenen op een eindejaarsuitkering, de zogenaamde 13e maand. SSC-I hecht sterk aan jouw persoonlijke groei en loopbaanontwikkeling en biedt daarvoor tal van mogelijkheden. Tot jouw secundaire arbeidsvoorwaarden behoren onder meer maximaal 55% betaald ouderschapsverlof (onder voorwaarden), studiefaciliteiten, een extra verlofregeling voor ouderen en een vergoeding woon-werkverkeer. Bij de Rijksoverheid heb je een aantal individuele keuzemogelijkheden bij het samenstellen van je arbeidsvoorwaardenpakket.

Bijzonderheden
Ben je in dienst van het Rijk en heb je een aanwijzing als VWNW- of (medisch) herplaatsing kandidaat? Voeg dan een voordracht van jouw leidinggevende of mobiliteitsadviseur én een kopie van de aanwijzingsbrief bij je sollicitatie.
Een assessment kan onderdeel uitmaken van de selectieprocedure.

Nadere informatie Meer informatie over de vacature kun je krijgen bij [NAME, FUNCTION, TELEPHONE NUMBER AND MAIL ADDRESS ANONYMISED BY SLM RIJK]

Meer informatie over de sollicitatieprocedure haal je bij Werving & Selectie [MAIL ADDRESS ANONYMISED BY SLM RIJK]

WORD DOCUMENT USECASE 6
Filename: Usecase6_original.doc

Existing Word document opened, the Connected Service used insert online picture, document stored in SharePoint Online

Opengevallen functie SSC-I

Sint Nicolaas (m/v)

Ministerie van Justitie & Veiligheid, Dienst Justitiële Inrichtingen
SSC-I, locatie Gouda (Schaal met pepernoten BBRA), 2 uur per jaar

Functieomschrijving
Het Shared Service Center ICT (SSC-I) in Gouda – een grote ICT-leverancier binnen de Overheid – is per 1-1-2018 op zoek naar een **SINT NICOLAAS.**

De kandidaat die wij zoeken, beschikt over natuurlijke senioriteit en engelengeduld. Een wilde, lange haardos en enorme baardgroei zijn een pré, maar geen must. Hij/zij werkt samen met afroep-Pieten en de leden van de Activiteitencommissie, die zich door de hele organisatie bevinden. De voorbereidende werkzaamheden vinden plaats in november/december van het jaar. De focus ligt hierbij op een twee uur durende aanwezigheid tijdens het jaarlijkse Sinterklaasfeest te Gouda, zulks ter bepaling door de Activiteitencommissie.

Functie-eisen
De persoon die wij zoeken, heeft een verantwoordelijke, veelomvattende functie: een groot aantal kinderen rekent op zijn/haar aanwezigheid, aandacht en cadeaus. De verantwoordelijkheden en taken bestaan uit het creëren van onvergetelijke herinneringen; het fungeren als aanspreekpunt voor alle kinderen van de medewerkers van SSC-I, de aansturing van een wisselend aantal Pieten uit verscheidene leeftijdscategorieën, het vullen van kinderschoenen en het vervullen van kinderwensen.

Voor het bijhouden van de administratie in het Grote Boek kan Sint Nicolaas rekenen op de permanente bijstand van collega's.

Omdat Sinterklaas een echte kindervriend is, beschikt hij/zij over een onuitputtelijke dosis aandacht en geduld. Verder is deze functionaris in het bezit van bovengemiddelde rijmkwaliteiten, goede paardrij-vaardigheden, het recept voor pepernoten en geen 9 tot 17 uur-mentaliteit.

Over SSC-I
SSC-I is dé ICT-dienstverlener van de Dienst Justitiële Inrichtingen van het ministerie van Veiligheid en Justitie. In totaal werken circa 450 vaste medewerkers bij SSC-I en 125 inhuur. Het merendeel daarvan werkt in Gouda, een ander deel in de nevenvestigingen in Zeist en Veenhuizen.
Buiten DJI bedienen wij diverse klanten binnen het veiligheidsdomein, zoals de Raad voor de Kinderbescherming (RvdK), de Dienst Terugkeer & Vertrek (DT&V), de Immigratie- en Naturalisatiedienst (IND) en Justis. Wij hebben de ambitie om de klantenkring binnen het ministerie van Veiligheid en Justitie verder uit te breiden.

Arbeidsvoorwaarden Salarisniveau: <niet van toepassing>
Salaris: ontelbare lachende en blije kindergezichtjes
Contractduur: vaste aanstelling (eventueel met een proeftijd)

Overige arbeidsvoorwaarden Bovenop het salaris kan Sinterklaas tijdens het uitoefenen van de functie rekenen op een gouden staf. Ook krijgt hij/zij werkkleding waaronder een mijter, tabberd en handschoenen. Paard: op aanvraag. Deze functie kent een aantrekkelijke bonusregeling, bestaande uit talloze kindertekeningen van verschillende kwaliteit en formaten, handjes, traantjes en ontzag.

Bijzonderheden
Een assessment door de leden van de Activiteitencommissie maakt deel uit van de selectieprocedure.

Nadere informatie Meer informatie over de vacature kun je krijgen bij vertrekkend Sint Nicolaas
[NAME, FUNCTION, TELEPHONE NUMBER AND MAIL ADDRESS ANONYMISED BY SLM RIJK]

Sollicitatie Direct solliciteren? Dat kan via deze link