Categories of private data

Cloud Manager

Tom Onacki
January 20, 2021

This PDF was generated from https://docs.netapp.com/us-en/occm/reference_private_data_categories.html on February 28, 2021. Always check docs.netapp.com for the latest.

Table of Contents

Categories of private data 1
- Types of personal data 1
- Types of sensitive personal data 3
- Types of categories 3
- Types of files 5
- Accuracy of information found 5

Categories of private data

There are many types of private data that Cloud Compliance can identify in your volumes, Amazon S3 buckets, databases, and OneDrive folders. See the categories below.

If you need Cloud Compliance to identify other private data types, such as additional national ID numbers or healthcare identifiers, email ng-contact-cloud-compliance@netapp.com with your request.

Types of personal data

The personal data found in files can be general personal data or national identifiers. The third column identifies whether Cloud Compliance uses proximity validation to validate its findings for the identifier.

Note that you can add to the list of personal data that is found in your files if you are scanning a database server. The *Data Fusion* feature allows you to choose the additional identifiers that Cloud Compliance will look for in its' scans by selecting columns in a database table. See Adding personal data identifiers using Data Fusion for details.

Type	Identifier	Proximity validation?
General	Email address	No
	Credit card number	No
	IBAN number (International Bank Account Number)	No
	IP address	No

Type	Identifier	Proximity validation?
National Identifiers	Belgian ID (Numero National)	Yes
	Brazilian ID (CPF)	Yes
	Bulgarian ID (UCN)	Yes
	California Driver's License	Yes
	Croatian ID (OIB)	Yes
	Cyprus Tax Identification Number (TIC)	Yes
	Czech/Slovak ID	Yes
	Danish ID (CPR)	Yes
	Dutch ID (BSN)	Yes
	Estonian ID	Yes
	Finnish ID (HETU)	Yes
	French Tax Identification Number (SPI)	Yes
	German Tax Identification Number (Steuerliche Identifikationsnummer)	Yes
	Greek ID	Yes
	Hungarian Tax Identification Number	Yes
	Irish ID (PPS)	Yes
	Israeli ID	Yes
	Italian Tax Identification Number	Yes
	Latvian ID	Yes
	Lithuanian ID	Yes
	Luxembourg ID	Yes
	Maltese ID	Yes
	Polish ID (PESEL)	Yes
	Portuguese Tax Identification Number (NIF)	Yes
	Romanian ID (CNP)	Yes
	Slovenian ID (EMSO)	Yes
	South African ID	Yes
	Spanish Tax Identification Number	Yes
	Swedish ID	Yes
	U.K. ID (NINO)	Yes
	USA Social Security Number (SSN)	Yes

Types of sensitive personal data

The sensitive personal data that Cloud Compliance can find in files includes the following:

Criminal Procedures Reference
Data concerning a natural person's criminal convictions and offenses.

Ethnicity Reference
Data concerning a natural person's racial or ethnic origin.

Health Reference
Data concerning a natural person's health.

ICD-9-CM Medical Codes
Codes used in the medical and health industry.

ICD-10-CM Medical Codes
Codes used in the medical and health industry.

Philosophical Beliefs Reference
Data concerning a natural person's philosophical beliefs.

Political Opinions Reference
Data concerning a natural person's political opinions.

Religious Beliefs Reference
Data concerning a natural person's religious beliefs.

Sex Life or Orientation Reference
Data concerning a natural person's sex life or sexual orientation.

Types of categories

Cloud Compliance categorizes your data as follows:

Finance

- Balance Sheets
- Purchase Orders
- Invoices
- Quarterly Reports

HR

- Background Checks
- Compensation Plans
- Employee Contracts
- Employee Reviews
- Health
- Resumes

Legal

- NDAs
- Vendor-Customer contracts

Marketing

- Campaigns
- Conferences

Operations

- Audit Reports

Sales

- Sales Orders

Services

- RFI
- RFP
- SOW
- Training

Support

- Complaints and Tickets

Metadata categories

- Application Data
- Archive Files
- Audio
- Business Application Data
- CAD Files
- Code
- Database and index files
- Design Files
- Email Application Data
- Executables
- Financial Application Data
- Health Application Data
- Images
- Logs
- Miscellaneous Documents
- Miscellaneous Presentations
- Miscellaneous Spreadsheets
- Videos

Types of files

Cloud Compliance scans all files for category and metadata insights and displays all file types in the file types section of the dashboard.

But when Cloud Compliance detects Personal Identifiable Information (PII), or when it performs a DSAR search, only the following file formats are supported:
.PDF, .DOCX, .DOC, .PPTX, .XLS, .XLSX, .CSV, .TXT, .RTF, and .JSON.

Accuracy of information found

NetApp can’t guarantee 100% accuracy of the personal data and sensitive personal data that Cloud Compliance identifies. You should always validate the information by reviewing the data.

Based on our testing, the table below shows the accuracy of the information that Cloud Compliance finds. We break it down by *precision* and *recall*:

Precision

The probability that what Cloud Compliance finds has been identified correctly. For example, a precision rate of 90% for personal data means that 9 out of 10 files identified as containing personal information, actually contain personal information. 1 out of 10 files would be a false positive.

Recall

The probability for Cloud Compliance to find what it should. For example, a recall rate of 70% for personal data means that Cloud Compliance can identify 7 out of 10 files that actually contain personal information in your organization. Cloud Compliance would miss 30% of the data and it won’t appear in the dashboard.

Cloud Compliance is in a Controlled Availability release and we are constantly improving the accuracy of our results. Those improvements will be automatically available in future Cloud Compliance releases.

Type	Precision	Recall
Personal data - General	90%-95%	60%-80%
Personal data - Country identifiers	30%-60%	40%-60%
Sensitive personal data	80%-95%	20%-30%
Categories	90%-97%	60%-80%

Copyright Information

Copyright © 2021 NetApp, Inc. All rights reserved. Printed in the U.S. No part of this document covered by copyright may be reproduced in any form or by any means-graphic, electronic, or mechanical, including photocopying, recording, taping, or storage in an electronic retrieval system-without prior written permission of the copyright owner.

Software derived from copyrighted NetApp material is subject to the following license and disclaimer:

THIS SOFTWARE IS PROVIDED BY NETAPP “AS IS” AND WITHOUT ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, WHICH ARE HEREBY DISCLAIMED. IN NO EVENT SHALL NETAPP BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

NetApp reserves the right to change any products described herein at any time, and without notice. NetApp assumes no responsibility or liability arising from the use of products described herein, except as expressly agreed to in writing by NetApp. The use or purchase of this product does not convey a license under any patent rights, trademark rights, or any other intellectual property rights of NetApp.

The product described in this manual may be protected by one or more U.S. patents, foreign patents, or pending applications.

RESTRICTED RIGHTS LEGEND: Use, duplication, or disclosure by the government is subject to restrictions as set forth in subparagraph (c)(1)(ii) of the Rights in Technical Data and Computer Software clause at DFARS 252.277-7103 (October 1988) and FAR 52-227-19 (June 1987).

Trademark Information

NETAPP, the NETAPP logo, and the marks listed at http://www.netapp.com/TM are trademarks of NetApp, Inc. Other company and product names may be trademarks of their respective owners.