

# Webroot Filters and Filter Logic

## Filter overviews for the U.S and Europe

### Introduction

Email Encryption Services powered by Zix use a set of comprehensive filters to scan for sensitive information, such as personal health information (PHI) or personal financial information (PFI) in electronic messages. Searches are conducted by scanning all email message subjects, bodies, and attachments for sensitive information defined within the filters.

A filter is a file consisting of a comprehensive set of terms, phrases, expressions and pattern masks that identify sensitive types of information. Sensitive information is defined as any information that, when inappropriately disclosed, can lead to significant contractual or legal liabilities; serious damage to your organization's image and reputation; or legal, financial, or business losses. Webroot uses many sources to generate the filter content that is used to search for sensitive information, including federal regulations, authoritative reference sources on the subject, and "standard of care" practices.

The following is a description of the filters that are typically used in Email Encryption Services, followed by a basic list of the formats inside each of the standard filters. In addition to these standard filters, custom filters can be created to detect sensitive information that is unique to an organization such as customer codes or classified project identifiers.

### United States Filters

#### Health Insurance Portability and Protection Act Regulation Filter (HIPAA)

The HIPAA regulatory filter is designed to identify PHI as defined by the Health Insurance Portability and Accountability Act (HIPAA). The HIPAA regulation filter contains a set of two filters; the HIPAA identifiers filter and the HIPAA health terms filter, that work together to identify PHI. The filters search for PHI by taking the intersection of identifying information, combined with health or claims information. This provides the highest level of confidence

that content is actually PHI. An example of this would be a document containing a patient's date of service and diagnosis. The date of service would constitute an identifier, and the diagnosis would constitute health information, thereby identifying PHI. The identifiers filter looks for indications of official business communications, such as medical record numbers, subscriber IDs, dates of birth, etc. The health terms filter scans for diagnoses, diseases, insurance information, pharmaceutical information, etc.

#### Health Insurance Claim Numbers and Medicare Beneficiary Identifier

In late 2018, the new Medicare Beneficiary Identifier (MBI) replaced the SSN-based Health Insurance Claim Number (HICN) on the new Medicare cards for Medicare transactions like billing, eligibility status and claim status. We recommend using both the Health Insurance Claim Number filter and the Medicare Beneficiary Number as outlined below.

To search for PHI, the Health Identifier and Health Terms filters are combined with the Health Insurance Claim filter, the Medicare beneficiary identifier filter and the Social Security Number filter using the following logic:

#### (HIPAA Identifiers AND HIPAA Terms) OR HIC number or MBI OR Social Security Number

- **Example #1:**

From: Sue  
To: Linda Subject:  
Subject: RE: Shared patient

Linda,  
Here's the info you requested on patient Jane Doe, **MRN 658578**. She sees Dr. A. at General Hospital. She began **fluorouracil** approximately 5/15/2019. When he saw her in 2020, he stated that she had been on **fluorouracil** for a year. Her last visit was 10/14/2020. No **cancer!**

- **Example #2:**

From: Sue  
To: Linda  
Subject: RE: Daily Inpatient Report

General Hospital does have an **acute rehab** service. Both members are improving considerably with their **therapy**. Members are Mr. Smith, **DOS 4/26/2020** & Mr. Jones, **DOS 5/11/2020**. They are on a **rehab** unit.

### Gramm, Leach, Bliley ACT (GLBA) Financial Regulation Filter

The Gramm, Leach, Bliley ACT (GLBA) Regulation filter consists of four filters: GLBA terms and GLBA identifiers, social security number and the Credit Card filter. These filters are designed to work in combination to recognize "Nonpublic Personally Identifiable Financial Information" as defined in the Gramm-Leach-Bliley Act (GLBA).

The filters work in conjunction to recognize the intersection of financial identifiers, such as

banking account numbers or loan numbers AND financial terms, such as "balance transfer", "refinance" or "deposit."

To search for PFI, the GLBA filters, the credit card filter and the social security number filter are combined using the following logic:

#### (GLBA identifiers AND GLBA Terms) OR Credit Card OR Social Security Number

The GLBA regulation filter can be used to assist companies in identifying personally identifiable financial information in email traffic.

Below are several example messages that would trigger the GLBA regulation filter. The expressions shown in bold font indicate terms that are identified in the filters.

- **Example #1:**

From: Linda  
To: Sue  
Subject: Account 12345

Dear Account holder,  
We here at Big-Mortgage-Finance Corp. have noticed that you have defaulted on **loan #123456**. We are happy to assist you however possible. Perhaps an automatic **payroll deduction** could help you make regular bill payments. Please see the attached **account summary** and submit payment in full as soon as possible to avoid **foreclosure**.

- **Example #2:**

From: Mike  
To: Daniel  
Subject: Prepayment Fees

In order to complete the monthly billing, please verify the **prepayment** fee for **the following accounts**:

- JOHN DOE 111001111 2,630.00  
- SUE JONES 222002222 4,250.00

Please respond as soon as possible, so we may complete the billing process. Thank you for your assistance.

### Financial Industry Regulatory Authority (FINRA)

Protection of financial and personal customer information is a key responsibility and obligation of FINRA member firms. This includes protecting against any anticipated threats or hazards to the security or integrity of customer records and information and against unauthorized access to, or use of customer records or information. The FINRA lexicon scans for 21 specific dealer and broker phrases along with extensive PII.

### Payment Card Industry Regulation Filter (PCI)

Major credit card companies and banks use standard numbering sequences that are unique to each brand of card, such as Visa, MasterCard or Discover. The Credit Card Number Filter can identify most credit card numbers and bank card numbers with matching technology that recognizes the identifiable patterns of numbers that all major credit card companies and banks use.

### Social Security Number Filter

This SSN filter is designed to identify social security numbers in emails. The filter is used to detect nine-digit numbers that meet the format requirements of an SSN or are found in close proximity to a label that identifies the number as an SSN. The SSN filter is included in every regulation filter as part of the policy-building process.

### Family Educational Rights and Privacy Act Regulation Filter (FERPA)

The Family Educational Rights and Privacy Act (FERPA) is a Federal law that protects the privacy of student education records. The FERPA filter scans for educational records including: student grade transcripts, student class schedules and student disciplinary records. Information that is deemed sensitive by this filter would be; social security numbers, grade point averages, student class schedules and student disciplinary reports.

## Personally Identifiable Information Filter (PII)

Many states and local municipalities have passed laws that protect the rights of an individual's personally identifiable information (PII). With that in mind, Webroot has created a PII filter that can be used to cover users from local and state laws. The PII filter scans for: social security numbers, driver's license numbers, dates of birth and many bank account numbers.

## Title Company Filter

In the mortgage industry, many types of sensitive information are passed through different channels. Financial data, credit card numbers and personally identifiable information are all transferred from mortgage companies to title companies regularly. With this in mind, Webroot created the title company filter.

This filter scans for SSN, credit cards, birthdates, bank account numbers and information, loan numbers, wire transfer information, driver's license numbers and more. This filter is not bound to any other filter and covers everything needed by a title company.

## Human Resource Filter

Human Resource departments routinely transmit sensitive employee data through email. While not specifically covered by any regulation most entities consider employee data sensitive. This filter was created to scan for sensitive employee information including; tax forms, pay stubs, employee rosters or other documents containing salary, or wage information.

## Sensitive Financial – Healthcare

The Sensitive Financial Healthcare Filters were created for healthcare entities that want to scan for sensitive financial information. This stand-alone filter comes in two variations; the standard version, which captures credit reports, banking account numbers and financial statements, and the expanded version which captures credit reports, banking account numbers, financial statements, wire information, ACH information and SWIFT codes.

## Profanity Filter

The profanity filter is designed to recognize profane and obscene language in email messages. According to Merriam-Webster, profane means "to debase by a wrong, unworthy, or vulgar use." Obscene means "marked by violation of accepted language inhibitions and by the use of words regarded as taboo in polite usage." These definitions form the basis on which this filter was designed and developed.

# United Kingdom / European Union

## United Kingdom - Personally Identifiable Information

The UK-Personally Identifiable Information (UK-PII) filter helps organizations identify personally identifiable information of citizens. The filter encompasses many different pattern masks to capture National Insurance Numbers, passport numbers, driver's license numbers, National Health Services numbers and dates of birth. This filter is recommended as the minimum email filtering a UK organization should deploy via Advanced Email Encryption.

## United Kingdom - Personal Health Information

The UK - Personal Health Information filters were created to mirror the U.S.-based HIPAA filters with language and pattern masks that are congruent with the UK. The filters search for PHI by taking the intersection of identifying information, combined with health or claims information. This provides the highest level of confidence that content is actually PHI. An example of this would be a document containing a patient's date of birth and diagnosis. The birthdate would constitute an identifier and the diagnosis would constitute health information. The identifiers filter looks for indications of official business communications, such as medical record numbers, dates of injury, dates of birth, etc. The health terms filter scans for diagnoses, diseases, pharmaceuticals, etc.

To search for PHI, the Health Identifier and Health Terms filters are using the following logic:

### (UK-PHI Identifiers AND UK-PHI Terms)

- **Example #1:**

From: Alan  
To: KATE  
Subject: Health information

Dear Sirs,  
RE: Julie Jones DOB 26/8/1982.

I refer to your correspondence received, 6th April, 2022, requesting Mental Health information for the above named.

I can confirm based on the information provided, that no, Mental Health records are held by, General Hospital Health Board, for, Julie Jones.

If you require any further information, please do not hesitate to contact me.

- **Example #2:**  
From: Bryan  
To: Megan  
Subject: RE: Referral

I have just been given a referral to see a Dermatologist via BUPA. Dr Bryan Ronald was one of the options suggested. Could you let me if it's possible to book a consultation and when is the next available slot? See my details below:

Name: David Graham  
DOB: 07/07/1981  
Address: 205 Stonemason House, 47 Hebb Road  
Post-code: E9 7HB  
Mobile: 07731548144  
Email: Dg@gmail.com  
Insurance:

- Membership: 0137955845
- Authorisation: 606165422

### United Kingdom - Personal Financial Information

Like the the UK – Personal Health Information filter, the UK – Personal Financial Information filter policy is a set of two filters which are designed to work together to recognize “Non-public Personally Identifiable Financial Information”. These two works in conjunction with the Credit Card filter to identify potentially sensitive financial information. The financial identifiers filter looks for many different bank account types and corresponding account numbers, pension plan accounts, sort codes, and swift codes. The financial terms lexicon looks for common banking terms like “deposit”, “bankruptcy” or “cash balance”. Together they find the intersection of common banking transactions found in email.

To search for PFI, the UK-PFI filters and the credit card filter are combined using the following logic:

#### (UK-PFI Identifiers AND UK-PFI Terms) OR Credit Card

- **Example #1:**  
From: David  
To: Vince  
Subject: Past due invoice

Dear Vince, Please find attached invoice INV-21946 for £662.40.

The balance amount outstanding of £331.20 is due by 14 Apr 2022.

Please check the invoice on receipt and advise of any queries.

Preferred forms of payment: BACS, Debit Card, Credit Card (except AMEX). Please use our invoice number as reference when paying by BACS.

Royal Bank of Scotland  
Sort Code 54-45-12  
Account No: 00457850

If you have any questions, please let us know.

- **Example #2:**  
From: Mike  
To: Daniel  
Subject: **Pension Payment**

Your September Pension Payment will be **deposited** into your **Current Account 00001** on 1/12/22.

If you need assistance, please call us at your convenience.

### EU General Data Protection Regulation (GDPR) Filter

The EU General Data Protection Regulation (GDPR) replaces the Data Protection Directive 95/46/EC and was designed to harmonize data privacy laws across Europe, to protect and empower all EU citizens' data privacy and to reshape the way organizations across the region approach data privacy. The GDPR regulation filter is a set of files that can be configured to be very broad or narrow depending on your approach. The narrowest versions capture all EU identifying numbers, health, financial, educational and employment information, while the broadest include health, financial and personal identification for all 27 EU countries.

### Customized Filters

The Webroot Research Center can help customers develop and deploy custom filters or design effective corporate email policies. For instance, a client may have specific account number or medical record formats, in this case the Webroot Research Center will create a filter to scan for those specific formats, thereby increasing the accuracy of that client's scanning capabilities. All customizations are performed as client service, and there is never any charge for this service. For instance, a client may have a specific account number or medical record format. Webroot will create a filter to scan for those specific formats, thereby increasing the accuracy of that client's scanning capabilities.

## Filter Development Process

Webroot goes to great lengths to ensure that filters are accurate and precise. This is accomplished through a comprehensive definition and design of the filters, coupled with exhaustive manual analysis to ensure that the filter results agree with the judgment of the filter designers. The following example provides a high-level overview of the design process and validation of the filters:

1. Standard filters designed based on definitions from HIPAA, GLBA, State Regulations or standard care practices.
2. Jury standard document developed
3. Message sample gathered from participating partner organizations
4. Samples manually examined using the jury standard document as a reference
5. Reference sources identified to ensure comprehensive content including medical dictionaries, professionally accepted terminology lists, legislation, etc.
6. Filters constructed and run against message samples
7. Filter results compared to manual results
8. Filters are tuned and re-run against the sample until performance is optimized
9. Revisions made based on changes to the definition of sensitive information and continuous collection of message samples.

## Content of Webroot Filters

The section below includes the basic information that each of the standard Webroot filters includes in its scanning formats and types of documents that each will scan for.

### HIPAA Identifiers

- Member Numbers
- Medical Savings Account numbers
- Medical Record Numbers
- Subscriber Numbers
- Patient ID numbers
- (All of the above are only found when in close proximity to a number at least 5 digits long)

- Admit Dates
- Dates of Birth
- Dates of Death
- Dates of Discharge
- Dates of Service
- Dates of Injury
- (All of the above are only found when in close proximity to a date)

### HIPAA Terms

- Diseases
- Chemicals
- Drugs
- Analytic, Diagnostic or Therapeutic Techniques
- Substance Use or Abuse
- Mental Health Terms
- Medical Records Information
- Insurance Information
- Medications

### GLBA Identifiers

- Vehicle Identification Numbers
- Banking Account Numbers (checking account, savings account, etc.)
- Certificate Numbers
- Loan Numbers
- Policy Numbers
- Member Numbers
- (All of the above are only found when in close proximity to a number at least 5 digits long)

### GLBA Terms

- Banking Terms
- Investment Terms
- Mortgage Terms
- General Financial Terms

### **FINRA Filter**

- Dealer/Broker phrases
- Banking terms
- Account numbers
- Dates of Birth
- Driver's license numbers
- Tax Identification numbers
- Social Security Numbers

### **Credit Card Number Filter**

- MasterCard formats
- Visa formats
- American Express formats
- Carte Blanche / Diners Club formats
- Discover formats
- Enroute formats
- JCB formats

### **Social Security Number Filter**

- Hyphenated 9-digit valid SSN sequence (###-##-####)
- 9-digit valid SSN sequence (#####) and in proximity of an SSN identifier (the phrase SSN, or SS, etc.)
- 9-digit valid SSN sequence separated by spaces or most any other punctuation (### ## #####) and in proximity of an SSN identifier (the phrase SSN, or SS, #, etc.)

### **Sensitive Financial –Healthcare Standard**

- Bank account numbers
- Budgeting Terms
- Credit Report information

### **Sensitive Financial –Healthcare Expanded**

- Bank account numbers
- Budgeting Terms
- Credit report information
- ACH Codes
- Wire Transfer Numbers/SWIFT Codes

### **Personally Identifiable Information**

- Bank Account Numbers
- Driver License formats
- Debit/Credit Card Numbers
- Birthdates (combined with place of birth, as seen on birth certificates)

### **Family Educational Rights and Privacy Act (FERPA)**

- Student Identification numbers
- Student Grade Point Averages
- Class lists/rosters
- Student transcripts

### **Human Resources**

- Job applications
- Employee rosters
- Resumes and CVs
- Disciplinary records
- Pay stubs

### **United Kingdom Personally Identifiable Information**

- National Insurance Numbers
- National Health Services Numbers
- Passport Numbers
- Driver License formats
- Birthdates

### **United Kingdom Personal Health Information Identifiers**

- Member Numbers
- Patient Numbers
- Medical Record Numbers
- Subscriber Numbers
- Patient ID Numbers
- National Health Service Numbers
- (All of the above are only found when in close proximity to a number at least 5 digits long)

- Admit Dates
- Dates of Birth
- Dates of Death
- Dates of Discharge
- Dates of Service
- Dates of Injury
- (All of the above are only found when in close proximity to a date)

### **United Kingdom Personal Health Information Terms**

- Diseases
- Chemicals
- Drugs
- Analytic, Diagnostic or Therapeutic Techniques
- Substance Use or Abuse
- Mental Health Terms
- Medical Records Information
- Medications

### **United Kingdom Personal Financial Information**

- Bank Account Numbers (standard accounts, savings accounts, etc.)
- Certificate Numbers
- Loan Numbers
- Policy Numbers
- Member Numbers
- (All of the above are only found when in close proximity to a number at least 5 digits long)
- Sort Codes
- Swift Codes

### **United Kingdom Personal Financial Terms**

- Banking Terms
- Investment Terms
- Mortgage Terms
- General Financial Terms

### **General Data Protection Regulation**

- European Union member's national identification numbers
- EU member's driver's license numbers
- EU member's tax identification numbers
- EU Health information including national insurance numbers
- EU Financial information including credit cards
- EU Educational information, including grade and disciplinary records
- EU Employment records including resumes, CVs and other employment records

#### **About Carbonite and Webroot**

Carbonite and Webroot, OpenText companies, harness the cloud and artificial intelligence to provide comprehensive cyber resilience solutions for businesses, individuals, and managed service providers. Cyber resilience means being able to stay up and running, even in the face of cyberattacks and data loss. That's why we've combined forces to provide endpoint protection, network protection, security awareness training, and data backup and disaster recovery solutions, as well as threat intelligence services used by market leading technology providers worldwide. Leveraging the power of machine learning to protect millions of businesses and individuals, we secure the connected world. Carbonite and Webroot operate globally across North America, Europe, Australia, and Asia. Discover cyber resilience at [carbonite.com](https://carbonite.com) and [webroot.com](https://webroot.com).

© 2022 OpenText. All rights reserved. OpenText, Carbonite, and Webroot are each trademarks of OpenText or its subsidiaries. All other trademarks are the properties of their respective owners. DS\_053122