

# Forcepoint

## Forcepoint Data Classification

Powered by Getvisibility

Data Classification Agent Metadata

**Forcepoint**

Report

# Table of Contents

- OVERVIEW .....2**
- METADATA ENTRIES .....2**
- AIP METADATA CONFIGURATION.....3**
- HOW TO GET MIP LABELS FROM AZURE.....4**
- CONFIGURABLE TAGS.....6**
- WRITING METADATA TO FILES .....7**
  - .ZIP .....7
  - .AVI, .WAV .....7
  - .MP4, .M4P, .M4V .....8
  - .MOV, .....8
  - .GIF, .JPG, .JPEG, .PNG, .TIFF, .TIF .....9
  - .VSDX, .DOCX, .XLSX, .XLSM, .PPTX .....9
  - .DOC, .XLS, .PPT .....10
  - .DXF, .DWG.....10
  - .PDF.....10



This design allows for a high degree of customization in how metadata is written, catering to diverse requirements and integration scenarios.

## AIP Metadata Configuration

The agent is able to write metadata that aligns with AIP's standards, allowing seamless interaction with Microsoft's security framework.

Below is an example of a configuration that demonstrates how the agent can be configured to write AIP metadata.

```
"aip": {
  "siteId": "ed86fd3a-ab24-4113-a9f8-6cb38f63c190",
  "labels": [
    {
      "id": "734f255c-faeb-4316-b0c0-3d88dfc5bbef",
      "name": "Public",
      "method": "Privileged",
      "contentBits": 3
    },
    {
      "id": "d2ee4459-2afa-4136-9d18-4f2ebca102cf",
      "name": "General Business",
      "method": "Privileged",
      "contentBits": 3
    },
    {
      "id": "65650f8c-b1ca-43b3-aad5-b4fa5ffcfb95",
      "name": "Confidential",
      "method": "Privileged",
      "contentBits": 3
    },
    {
      "id": "574e5463-b443-4182-9781-3fc620ed259c",
      "name": "Restricted",
      "method": "Privileged",
      "contentBits": 3
    }
  ],
  "mappings": [
    {
      "classification": "Public",
      "aipLabelId": "734f255c-faeb-4316-b0c0-3d88dfc5bbef"
    },
    {
      "classification": "General Business",
      "aipLabelId": "d2ee4459-2afa-4136-9d18-4f2ebca102cf"
    },
    {
      "classification": "Confidential",
      "aipLabelId": "65650f8c-b1ca-43b3-aad5-b4fa5ffcfb95"
    },
    {
      "classification": "Restricted",
      "aipLabelId": "574e5463-b443-4182-9781-3fc620ed259c"
    }
  ]
},
```

This configuration includes specific AIP labels and their corresponding classifications, defining how the agent translates its internal categorizations into a format that AIP can recognize.

# How to get MIP labels from Azure

We can get the labels configured in customer's Azure using the Fetch MIP labels from Azure in power tools.

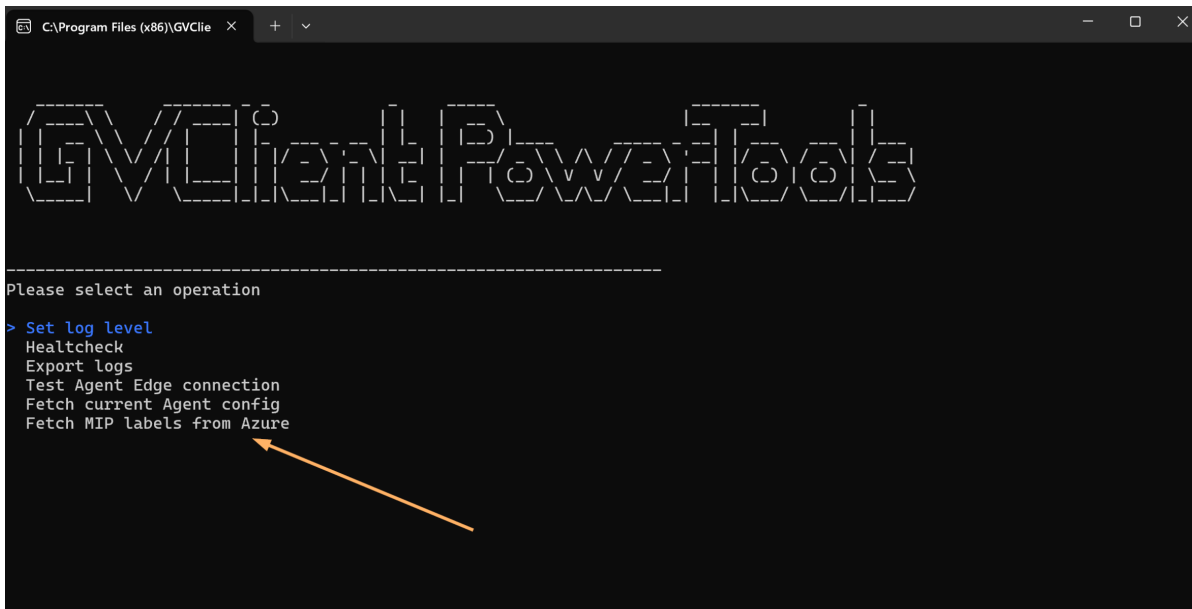


figure 1.

It accepts all the required details from the user (tenantId, appName, clientId, clientSecret, emailId) as inputs and prints out the available MIP labels as output.

```

Total MIP Labels found: 5
The MIP labels:
{
  "siteId": "ed86fd3a-ab24-4113-a9f8-6cb38f63c190",
  "labels": [
    {
      "id": "734f255c-faeb-4316-b0c0-3d88dfc5bbef",
      "name": "Public",
      "method": "Priviledged",
      "contentBits": 3
    },
    {
      "id": "d2ee4459-2afa-4136-9d18-4f2ebca102cf",
      "name": "General",
      "method": "Priviledged",
      "contentBits": 3
    },
    {
      "id": "65650f8c-b1ca-43b3-aad5-b4fa5ffcfb95",
      "name": "Confidential",
      "method": "Priviledged",
      "contentBits": 3
    },
    {
      "id": "574e5463-b443-4182-9781-3fc620ed259c",
      "name": "Highly Confidential",
      "method": "Priviledged",
      "contentBits": 3
    },
    {
      "id": "4ceb0376-6028-4ea5-abd6-f8cb9ce4033c",
      "name": "Compliance",
      "method": "Priviledged",
      "contentBits": 3
    }
  ]
}
Press any key to continue... (continue):

```

figure 2.

The output format:

```

{
  "siteId": "tenant_id",
  "labels": [
    {
      "id": "label_id",
      "name": "label_name",
      "method": "label_assignment_method", // always Priviledged
      "contentBits": type of content marking applied // always 3
    }
  ]
}

```

Reference: [Concepts - Label metadata in the MIP SDK](#)

This utility will use application permissions in azure app and the following API permissions need to be provided to it (with admin

consent):

- Microsoft Graph → InformationProtectionPolicy.Read.All
- Microsoft Information Protection Sync Service → UnifiedPolicy.Tenant.Read

## Configurable Tags

The agent also provides the functionality to write fully customized metadata entries.

Below is an example of a configuration that defines a variety of custom metadata tags:

```
{
  "id": "8e808d02-fdc4-4fc9-8938-853b0a4a5d3d",
  ...
  "defaultSettings": {
    ....
    "writeMetadataTags": [
      {
        "tagName": "ClassificationTagSetId",
        "tagHandle": "e16409a7-1700-4153-9090-3955bc2f0ae8"
      },
      {
        "tagName": "ClassificationValue",
        "tagHandle": "{classification}"
      },
      {
        "tagName": "DistributionTagSetId",
        "tagHandle": "004dea33-8751-4399-a76e-95f371cb4119"
      },
      {
        "tagName": "DistributionValue",
        "tagHandle": "{distribution}"
      },
      {
        "tagName": "FileId",
        "tagHandle": "{fileid}"
      },
      {
        "tagName": "UserId",
        "tagHandle": "{user}"
      },
      {
        "tagName": "TagDateTime",
        "tagHandle": "{datetime}"
      },
      {
        "tagName": "{name:e16409a7-1700-4153-9090-3955bc2f0ae8:0}",
        "tagHandle": "{value:e16409a7-1700-4153-9090-3955bc2f0ae8:0}"
      },
      {
        "tagName": "classification_guid",
        "tagHandle": "{classification_guid}"
      },
    ],
    {
      "tagName": "compliance_guid",
      "tagHandle": "{compliance_guid}"
    },
  ],
  {
    "tagName": "distribution_guid",
```

```

      "tagHandle": "{distribution_guid}"
    }
  ],
},
...
}

```

The `tagHandle` field within this configuration supports various placeholders that facilitate dynamic tagging:

- `{classification}` - current classification value of the document
- `{distribution}` - current distribution value of the document
- `{compliance}` - current compliance value of the document
- `{datetime}` - current date and time
- `{email}` - email of the current user (only works when outlook plugin is installed)
- `{user}` - id of the current user
- `{machineid}` - id of the current machine
- `{fileid}` - unique file id
- `{classification_raw}` - This is the tag value without the tag alias
- `{compliance_raw}` - This is the tag value without the tag alias
- `{distribution_raw}` - This is the tag value without the tag alias
- `{classification_guid}` - This is a unique id generated based on tagset id and tag name
- `{classification_guid}` - This is a unique id generated based on tagset id and tag name
- `{classification_guid}` - This is a unique id generated based on tagset id and tag name

## Writing Metadata to files

As already mentioned in the beginning of this page, the agent uses a slightly different approach for each file to write/read the metadata. This flexibility is necessitated by the diverse range of file types and their various mechanisms for storing metadata. The following outlines the approaches used for different file categories:

### **.zip**

For `.zip` files, the agent creates a file named `GV_metadata.json` within the classified zip file. The format is analogous to the example provided earlier.

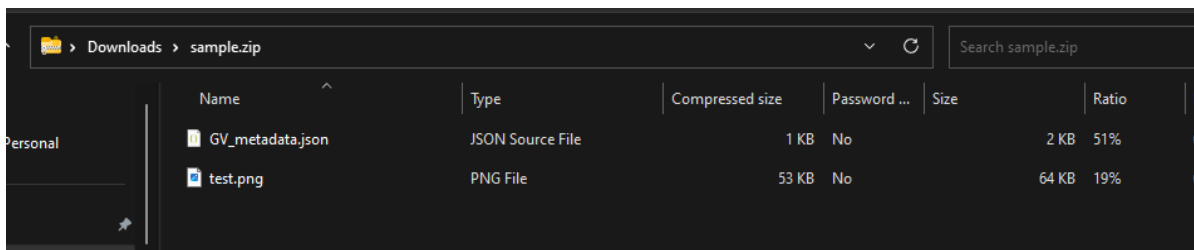


figure 3.

### **.avi, .wav**

D3v2 tags are utilized for `.avi` and `.wav` files. These tags are variable in size and strategically placed at the beginning of the file,



facilitating immediate metadata loading, even during incremental file streaming. An ID3v2 tag comprises multiple optional frames, each encapsulating metadata up to 16 MB in size.

The written metadata can be accessed via the tag `comment-pol`.

warning	Invalid ID3 frame size
comment-pol	{ "GVData": "ew0KICAidGFnc2V0X2YxNGZlMwYxZg5NTBfNDBkNV84YTl5XzQ1OTA5ZG5ESNdDkNl9nZHBYL3BpaSI6lCjYVWxzSl5DQogICJ0YVdzZXRfZjE0ZmMxZjFfODk1MF80MGQ1Xzhh", "GVData0": "MDA0ZGVhMzNfODc1MV80Mzk5X2E3NmVfOTVmMzcY2l0MTE5X2Rpc3RyaWJ1dGlvbll6lCjbnRlcm5hbCIsDQogICJ0YVdzZXRfZTE2NDAsYTDfMTcwMF80MTUzXzkvOTBfMzk1", "GVData1": "(end)", "ClassificationTagSetId": "e16409a7-1700-4153-9090-3955bc2f0ae8", "ClassificationValue": "General Business", "DistributionTagSetId": "004dea33-8751-4399-a76e-95f371cb4119", "DistributionValue": "Internal / u0645u0631u0628u0627u0628u0627u0644u0639u0627u0644u0645", "Field": "75adb5db-35a6-4f52-9244-3903a01f62f7", "UserId": "trzec", "TagDateTime": "2023-08-02T14:41:52Z", "Classification": "General Business", "MSIP_Label_d2ee4459-2afa-4136-9d18-4f2ebca102cf_ActionId": "582d90ff-586d-4ae4-97ad-c8340eb69ade", "MSIP_Label_d2ee4459-2afa-4136-9d18-4f2ebca102cf_ContentBits": "3", "MSIP_Label_d2ee4459-2afa-4136-9d18-4f2ebca102cf_Enabled": "true", "MSIP_Label_d2ee4459-2afa-4136-9d18-4f2ebca102cf_Method": "Privileged", "MSIP_Label_d2ee4459-2afa-4136-9d18-4f2ebca102cf_Name": "General Business", "MSIP_Label_d2ee4459-2afa-4136-9d18-4f2ebca102cf_SetDate": "2023-08-02T14:41:52Z", "MSIP_Label_d2ee4459-2afa-4136-9d18-4f2ebca102cf_SiteId": "ed86fd3a-ab24-4113-a9f8-6cb38f63c190" }
software	Lavf57.19.100

figure 4.

**.mp4, .m4p, .m4v**

These file types make use of the ILST tag format. Further details can be found on the [Apple metadata page](#).

The written metadata can be accessed via the tag `comment`.

creation_time	2015-08-07T09:13:02.000000Z
comment	{ "GVData": "ew0KICAidGFnc2V0X2YxNGZlMwYxZg5NTBfNDBkNV84YTl5XzQ1OTA5ZG5ESNdDkNl9nZHBYL3BpaSI6lCjYVWxzSl5DQogICJ0YVdzZXRfZjE0ZmMxZjFfODk1MF80MGQ1Xzhh", "GVData0": "MDA0ZGVhMzNfODc1MV80Mzk5X2E3NmVfOTVmMzcY2l0MTE5X2Rpc3RyaWJ1dGlvbll6lCjbnRlcm5hbCIsDQogICJ0YVdzZXRfZTE2NDAsYTDfMTcwMF80MTUzXzkvOTBfMzk1", "GVData1": "(end)", "ClassificationTagSetId": "e16409a7-1700-4153-9090-3955bc2f0ae8", "ClassificationValue": "General Business", "DistributionTagSetId": "004dea33-8751-4399-a76e-95f371cb4119", "DistributionValue": "Internal / u0645u0631u0628u0627u0628u0627u0644u0639u0627u0644u0645", "Field": "b2b718ad-869b-a6ad-9fcd-72748bab65b0", "UserId": "trzec", "TagDateTime": "2023-08-02T14:31:00Z", "Classification": "General Business", "MSIP_Label_d2ee4459-2afa-4136-9d18-4f2ebca102cf_ActionId": "6d6af634-bf54-4eea-918d-c702dd8cc627", "MSIP_Label_d2ee4459-2afa-4136-9d18-4f2ebca102cf_ContentBits": "3", "MSIP_Label_d2ee4459-2afa-4136-9d18-4f2ebca102cf_Enabled": "true", "MSIP_Label_d2ee4459-2afa-4136-9d18-4f2ebca102cf_Method": "Privileged", "MSIP_Label_d2ee4459-2afa-4136-9d18-4f2ebca102cf_Name": "General Business", "MSIP_Label_d2ee4459-2afa-4136-9d18-4f2ebca102cf_SetDate": "2023-08-02T14:31:00Z", "MSIP_Label_d2ee4459-2afa-4136-9d18-4f2ebca102cf_SiteId": "ed86fd3a-ab24-4113-a9f8-6cb38f63c190" }

figure 5.

**.mov,**

These files leverage XMP tags. XMP, or Extensible Metadata Platform, is an Adobe-created standard for embedding metadata within digital files. This approach ensures standardized, structured metadata embedding that is readily extendable, encompassing information such as creator, copyright, editing data, and more.

The written metadata can be accessed via the tag `classification`.

xmp_toolkit	Image:ExifTool 12.44
classification	{ "GVData": "ew0KICAlidGFnc2V0X2YxNGZjMwYxXz5NTBfNBkNv84YTI5xZQ1OTA5ZGE5NDdkNl9hZHB5L3BpaS16lCjYwXzZSIsDQogICJ0YWdzZXRfZTE2NDAsYTFdfMTcwMF80MTUzXzkWOTBfMzkz "GVData0": "MDA0ZGVhMzNfODc1Mv80Mzk5X2E3NmVfOTVhMzcxY2l0MTE5X2Rpc3RyaWJ1dGlvbll6lCjBnRlcmShbC1sDQogICJ0YWdzZXRfZTE2NDAsYTFdfMTcwMF80MTUzXzkWOTBfMzkz "GVData1": "ICAIZG9jSUQlQlA2U040WmYjATNGFHNC1hYTQ4LTg0YzltYzMyOWE4Y2Q4M2ZmlwNCiAgIk9TjogldpbmRvd3MIDQp9", "GVData2": "(end)", "ClassificationTagSetId": "e16409a7-1700-4153-9090-3955bc2f0ae8", "ClassificationValue": "General Business", "DistributionTagSetId": "004dea33-8751-4399-a76e-95f371cb4119", "DistributionValue": "Internal / u0645u0631u062Du0628u0627 u0628u0627u0644u0639u0627u0644u0645", "FileId": "7e89efb0-4aa4-aa48-84c2-c329a8cd83ff", "UserId": "trzec", "TagDateTime": "2023-08-02T14:27:56Z", "Classification": "General Business", "MSIP_Label_d2ee4459-2afa-4136-9d18-4f2ebca102cf_ActionId": "85f23af2-7b8a-455c-9041-f59b5391b360", "MSIP_Label_d2ee4459-2afa-4136-9d18-4f2ebca102cf_ContentBits": "3", "MSIP_Label_d2ee4459-2afa-4136-9d18-4f2ebca102cf_Enabled": "true", "MSIP_Label_d2ee4459-2afa-4136-9d18-4f2ebca102cf_Method": "Privileged", "MSIP_Label_d2ee4459-2afa-4136-9d18-4f2ebca102cf_Name": "General Business", "MSIP_Label_d2ee4459-2afa-4136-9d18-4f2ebca102cf_SetDate": "2023-08-02T14:27:56Z", "MSIP_Label_d2ee4459-2afa-4136-9d18-4f2ebca102cf_SiteId": "ed86fd3a-ab24-4113-a9f8-6cb38f63c190" }
media_data_size	675191

figure 6.

**.gif, .jpg, .jpeg, .png, .tiff, .tif**

These files also leverage XMP tags to write metadata. The written metadata can be accessed via the tags `description` and `user_comment`.

interlace	Noninterlaced
description	{ "GVData": "ew0KICAlidGFnc2V0X2YxNGZjMwYxXz5NTBfNBkNv84YTI5xZQ1OTA5ZGE5NDdkNl9hZHB5L3BpaS16lCjYwXzZSIsDQogICJ0YWdzZXRfZTE2NDAsYTFdfMTcwMF80MTUzXzkWOTBfMzkz "GVData0": "MDA0ZGVhMzNfODc1Mv80Mzk5X2E3NmVfOTVhMzcxY2l0MTE5X2Rpc3RyaWJ1dGlvbll6lCjBnRlcmShbC1sDQogICJ0YWdzZXRfZTE2NDAsYTFdfMTcwMF80MTUzXzkWOTBfMzkz "GVData1": "(end)", "ClassificationTagSetId": "e16409a7-1700-4153-9090-3955bc2f0ae8", "ClassificationValue": "General Business", "DistributionTagSetId": "004dea33-8751-4399-a76e-95f371cb4119", "DistributionValue": "Internal / u0645u0631u062Du0628u0627 u0628u0627u0644u0639u0627u0644u0645", "FileId": "e0481ca0-a9e0-e307-07fa-6189581762a8", "UserId": "trzec", "TagDateTime": "2023-08-02T11:17:28Z", "Classification": "General Business", "MSIP_Label_d2ee4459-2afa-4136-9d18-4f2ebca102cf_ActionId": "95871ebc-c143-40b9-9b42-ad7bd6bc77df", "MSIP_Label_d2ee4459-2afa-4136-9d18-4f2ebca102cf_ContentBits": "3", "MSIP_Label_d2ee4459-2afa-4136-9d18-4f2ebca102cf_Enabled": "true", "MSIP_Label_d2ee4459-2afa-4136-9d18-4f2ebca102cf_Method": "Privileged", "MSIP_Label_d2ee4459-2afa-4136-9d18-4f2ebca102cf_Name": "General Business", "MSIP_Label_d2ee4459-2afa-4136-9d18-4f2ebca102cf_SetDate": "2023-08-02T11:17:28Z", "MSIP_Label_d2ee4459-2afa-4136-9d18-4f2ebca102cf_SiteId": "ed86fd3a-ab24-4113-a9f8-6cb38f63c190" }
user_comment	{ "GVData": "ew0KICAlidGFnc2V0X2YxNGZjMwYxXz5NTBfNBkNv84YTI5xZQ1OTA5ZGE5NDdkNl9hZHB5L3BpaS16lCjYwXzZSIsDQogICJ0YWdzZXRfZTE2NDAsYTFdfMTcwMF80MTUzXzkWOTBfMzkz "GVData0": "MDA0ZGVhMzNfODc1Mv80Mzk5X2E3NmVfOTVhMzcxY2l0MTE5X2Rpc3RyaWJ1dGlvbll6lCjBnRlcmShbC1sDQogICJ0YWdzZXRfZTE2NDAsYTFdfMTcwMF80MTUzXzkWOTBfMzkz "GVData1": "(end)", "ClassificationTagSetId": "e16409a7-1700-4153-9090-3955bc2f0ae8", "ClassificationValue": "General Business", "DistributionTagSetId": "004dea33-8751-4399-a76e-95f371cb4119", "DistributionValue": "Internal / u0645u0631u062Du0628u0627 u0628u0627u0644u0639u0627u0644u0645", "FileId": "e0481ca0-a9e0-e307-07fa-6189581762a8", "UserId": "trzec", "TagDateTime": "2023-08-02T11:17:28Z", "Classification": "General Business", "MSIP_Label_d2ee4459-2afa-4136-9d18-4f2ebca102cf_ActionId": "95871ebc-c143-40b9-9b42-ad7bd6bc77df", "MSIP_Label_d2ee4459-2afa-4136-9d18-4f2ebca102cf_ContentBits": "3", "MSIP_Label_d2ee4459-2afa-4136-9d18-4f2ebca102cf_Enabled": "true", "MSIP_Label_d2ee4459-2afa-4136-9d18-4f2ebca102cf_Method": "Privileged", "MSIP_Label_d2ee4459-2afa-4136-9d18-4f2ebca102cf_Name": "General Business", "MSIP_Label_d2ee4459-2afa-4136-9d18-4f2ebca102cf_SetDate": "2023-08-02T11:17:28Z", "MSIP_Label_d2ee4459-2afa-4136-9d18-4f2ebca102cf_SiteId": "ed86fd3a-ab24-4113-a9f8-6cb38f63c190" }
srgb_rendering	Perceptual

figure 7.

**.vsdx, .docx, .xlsx, .xlsm, .pptx**

Microsoft Office allows for Custom Properties to be written to these documents, stored internally in XML format. The metadata will be written as separate key/values tag pairs instead of JSON.



gv_data	ew0KICAidGfnc2V0X2YxNGZlMwYxXzgz5NTBfNDBkNV84YTI5XzQ1OTA5ZGE5NDdkNl9nZHBzL3BpaS6lCjYwXzZ5isDQogICJ0YWdzZXRfZjE0ZmMxZjFODk1MF80MGQ1XzhM
gv_data0	MDA0ZGVhMzNfODc1MV80Mzk5X2E3NmVfOTVmMzcY2l0MTE5X2Rpc3RyaWJ1dGlvbiI6lCjJbnRlcm5hbC1sDQogICJ0YWdzZXRfZTE2NDAsYVtdfMTcwMF80MTUzXzkwOTBfMzk
gv_data1	(end)
classification_tag_set_id	e16409a7-1700-4153-9090-3955bc2f0ae8
classification_value	General Business
distribution_tag_set_id	004dea33-8751-4399-a76e-95f371cb4119
distribution_value	Internal / مرصدا بلعلم
file_id	82b0ca65-575e-8f3e-cd97-1d372e7fe17f
user_id	trzec
tag_date_time	2023-08-02T14:46:08Z
classification	General Business

figure 10.



[forcepoint.com/contact](https://forcepoint.com/contact)

## About Forcepoint

Forcepoint is the leading user and data protection cybersecurity company, entrusted to safeguard organizations while driving digital transformation and growth. Forcepoint's humanly-attuned solutions adapt in real-time to how people interact with data, providing secure access while enabling employees to create value. Based in Austin, Texas, Forcepoint creates safe, trusted environments for thousands of customers worldwide.