foldingspace
Regent House, 70 Caroline Street, Jewellery Quarter, Birmingham B3 1UG
Tel: 0121-236-8979 Fax: 0121-236-9035
studio@foldingspace.co.uk www.foldingspace.co.uk

Folding Space Automated Discovery & Detection (ADD) Service

SERVICE DESCRIPTION

ADD - Solution Summary

As an innovative finding technology, ADD discovery provides the automated, programmatic means to answer all the key questions about all forms of unstructured data.

ADD discovery derives from a unique collection of technologies, techniques, tools and original software developments that power the various discovery processes.

- Universal file detection and document recognition
- High-precision OCR with High-performance text identification
- Exceptionally intricate content perception & extraction
- Very powerful & hugely scalable data indexing

ADD - Service Description

ADD can be provided as a one-off service or as an ongoing analysis of data to support several different use cases described later in this document. Here are the core Discovery processes (Adoption of at least File & Document Discovery is a prerequisite for all customers):

File & Document Discovery

Find, analyse, report and monitor all files & documents across all customer file stores wherever they are located.

Access Discovery (including Microsoft Active Directory (MAD) Lookup)

Identify who has access to the found documents, images, files & records.

Content Discovery

Find, extract, recognise & report on all the text inside the found documents.

Database Discovery

Find whatever metadata is relevant for whatever purpose from within multiple data streams.

PDF Discovery

PDF content & metadata discovery, document detection & separation, file conversion & formatting.

In complement to Discovery, Detection is an analytic technology that provides the analyses & interpretation of what has been discovered.

Detection derives from highly advanced IT/AI technologies & techniques that are state-of-the-art and beyond; for example: -

- Uniquely extensive & customisable data dictionaries
- Sophisticated libraries of complex, nested regular expressions
- Detailed text pattern, cluster, context & proximity algorithms

- AI (Artificial Intelligence) with a particular focus on very advanced areas of NLP (Natural Language Processing) like NED (Named Entity Disambiguation)

Automated Discovery & Detection (ADD) combines discovery & detection technologies into a coherent and complementary software platform that can be customised, configured and targeted at any and many file stores and document corpus to suit the customer purpose. Equally, ADD leads on to a variety of applications, such as:

- Document deduplication
- Email deduplication & threading
- Image deduplication
- Metadata extraction
- PDF metadata extraction, document detection & splitting
- Personal Data discovery (GDPR)
- ROT management
- Sensitivity analysis
- User/Document permissions analysis

ADD application Illustrations

Metadata extraction: ADD can uncover the facts, figures, dates, terms, values, labels, descriptors and similar currently hidden within document text. And provide automated identification, analysis & extraction of relevant data from within any volume of documents to drive actions, insight, processes and workflows.

Personal Data (GDPR) discovery: ADD can provide the most automated, accurate, practical and secure solution for finding, viewing, monitoring, & reporting upon the existence, occurrence and processing of Personal Data within every document anywhere across the organisation irrespective of network location, format or storage.

PDF analysis & retrieval: ADD can automate the content indexing, semantic analysis and splitting apart of large PDF files into their constituent documents. Typically, this occurs when paper records like personnel files have been scanned into single PDFs and now need to be separated for data retention and privacy compliance. Equally, ADD can collate & convert multiple documents of any format into PDF files.

ROT: Redundant, Obsolete & Trivial file management

Making ROT manageable, reducing risk, saving money

Whilst ADD discovers all the files in all the target file stores, ROT then addresses the issue of detection & remediation of files that are redundant, obsolete or trivial.

> **Redundant** refers to unnecessary duplicates - copies of files stored in multiple places and across multiple systems – shared drives are the most common containers of document duplicates whilst copies of emails are the most prevalent form of file duplication

Obsolete because the files are incorrect, incomplete or simply no longer in use. Like outdated document content that has been superseded by new information, or a PowerPoint or Excel or similar from 10 years ago or files generated by defunct programs. Effectively, any file that is 'old' or dated by current business standards or processes and that does not need to be kept for future reference, record-keeping, regulation or legal requirement.

Trivial because there is no need to store these files – they don't need to be kept as they do not contribute to corporate knowledge, business insight or record-keeping requirements. Typically, these 'junk files' are temporary or transitory. There is no value to the organisation, for instance, in files with extensions such as .001, .chk, .gid, .prv, .tmp, nor in out-of-office emails and similar; but they clog up storage and can be a security risk (e.g. an open invitation to hackers).

So, ADD first finds, analyses, reports and monitors each and every file, document and record across the customer network and/or cloud estate - wherever they exist in servers, storage, drives, directories, folders and archives. Then, automated ROT processes quantify & qualify each and every type of file/document/record as being Redundant, Obsolete or Trivial (or not).

ROT can also provide direct access to each file irrespective of file location. And ROT can then enable the automated disposition of each ROT file (archive, delete, keep, move, quarantine). Essentially, ROT provides the ability to set actions and workflows to deal with the ROT files found by Discovery & Detection.

A policy of ROT file definitions, rules & dispositions is agreed with each customer; typically, this is driven by the application of a 'best practice' Blueprint. This approach increases the speed at which an effective ROT policy can be implemented.

Beyond, the key metadata attributes of each & every folder and file discovered per file container (location, pathway, name, type, extension, size, created, modified, accessed, etc.), ROT extends into quantifying and qualifying the file extensions of all the discovered files as this informs the discovery interpretation and can immediately help to uncover ROT. In particular, Obsolete and Trivial files.

Usefully (and unusually), Folding Space created and maintains a database of all File Extensions currently comprising 10 categories, 30 sub-categories and 11,129 file extensions.

The key point, of course, is that if all file extensions can be found and then interpreted so that their 'parent' program is identified and/or their origin or nature is determined, then their status as ROT (or not) can be better gauged.

All the results generated by the various ROT processes are analysed, interpreted and reported upon automatically so customers can see at a glance exactly what they are dealing with in terms of ROT occurrence and instances. ROT can be applied as a one-off project to clean out ROT files but, more usefully, it can be a continuing process to guard against the recurrence of ROT.

Consequently, with ROT, there can be a regularly renewing measure of meaningful output from all the ROT processes so that Analysis, Reporting & Visualisation are continually presented. Because all ROT candidates discovered are defined in detail per file container and are categorised as per the ROT Policy/Blueprint. All of these granular results are automatically aggregated (with drill-down and drill-up facilities) into illuminating charts and illustrative visualisations for easy reference and an overview of the ROT situation.

The ROT Policy/Blueprint also answers the question as to how to address and deal with the ROT (and not ROT) files. Because ROT can also enable the automated disposition of discovered files according to the agreed ROT Policy/Blueprint. Typically, the ROT Policy/Blueprint involves a choice of file disposition – to archive (i.e. migrate to lower cost storage, move offline, etc.) – to delete (ensuring it is safe and defensible to do so) - to keep (the file is not ROT) - to move (the file is in the wrong container/location) - or to quarantine (awaiting confirmation to delete from authorised users). The automated dispositions (action rules) have customer agreed safety checks and a full audit of actions is set in parallel.

In summary; ROT finds, analyses, reports (and can continually monitor) each & every file, document & record across an organisation's network and their cloud estate to uncover the ROT so it can be dealt with before it becomes a bigger problem than already exists. This enables the organisation to address the cost, misinformation, productivity, security and liability losses caused by ROT immediately, and going forward.

Core Service Offering

The service is configured and operated via the following model. Firstly, an on premises physical or virtual file servers for the file finding and scanning service (DHS), secondly from an environment hosted in our secure private cloud (ADD) to deliver reports and results, all delivered from our Tier III, ISO accredited data centre facility. It presents the ability to provide for user data (unstructured data) management with an option to remediate data either to delete, move or archive, based on the findings from the scan.

Service Connectivity

The environment requires the provision of a secure VPN tunnel between the customer site and the service data centre, to enable the service. This will connect the local (customer) network to our private cloud network and a VPN will be established to secure the connection. The exact hardware configuration required to achieve this (dependent upon the customer requirement) will be part of the consultation process at service take-on and is a precursor for the service itself.

Service Options

There are also options (not included in the core service as standard) to include an-ongoing scanning service (delivered at a quarterly frequency), remediation services which requires a contract to the data governance service and data consultancy as part of the core service, as required. Service options are designed to support and enhance the core solution and are optional.

Service Specific Consultancy

As a pre-cursor to service, we can provide consultancy as to the best solution options for your requirements and assist you in deploying and configuring the solution options you require.

Service Availability & Redundancy

Information Assurance

All parties providing cloud services hold at least ISO/IEC27001:2013 and Cyber Essentials. As required, we will use appropriately certified and accredited staff with appropriate recognised processes (e.g. ITIL, Prince etc.) and clearance to complete the necessary project.

Service on-boarding

New customers will require a project to scope, design and define the sources that we will be dealing with, and the availability of computing resources on site (either physical or virtual) to provide the scanning from our DHS engine. Any project work will identify the required technical and cost components from a Cloud perspective, and the required effort costed on a T&M basis using our current day rates for the work being undertaken.

Service off boarding

Customers exiting this service will require a project to identify and manage the closure of the cloud service, the recovery of any data stored on the cloud platforms. The exact scope of work will depend upon the nature of the service being exited, but an Off Boarding plan will form a contractual part of any Service Take-On, ensuring responsibilities are clear and defined for peace of mind.

Customer Responsibilities

The following are the main responsibilities undertaken by the customer in any given cloud service scenario: -

- Acceptance of any Cloud provider Agreements and/or Terms and Conditions for service access and elements provided by Private Cloud Services Agreement, AWS and Microsoft
- Payment of all cloud service and incremental charges required by the cloud provider connected to the running of the service
- Acceptance of on-boarding and off-boarding project costs
- Compliance with the terms of any communications agreement for connectivity services, including installation and termination charges where appropriate

Service Management

We are a service management focused organisation with experience of service management implementation, operation and we implement continuous service improvement policies. Our operational processes are aligned with ITIL V3 methodologies, and we are also ISO/IEC27001:2013 accredited for information assurance.

Support

Our service is supported by our Service Desk. We operate a full range of standard support services between the hours of 09:00 and 17:00.

Incident Management

Our support services include Incident Management and triage services for all Cloud Services. We use our service processes to log, assign and diagnose incidents and to restore services with a minimum of disruption in line with our agreed hours of service.

Incidents are managed Monday to Friday 09:00 to 17:00, unless additional support hours have been contracted. If necessary, we will contact the Cloud service provider (Amazon or Microsoft) should our triage process find service issues relating to them. Incidents relating to major service failure in the Public Cloud (if adopted as part of the solution) are not included in our targets, as they are out of our direct control.

Severity Level P1
Service completely unavailable. Service Level Target miss (or expected miss) Response Critical system or component failure where no failover - updates to problem record hourly or upon change of status, work 24 x 7 until problem is resolved Target resolution: 4 Support Hours

Severity Level P2
Service degraded (e.g. slow response times). Non-critical system or component failure Response Updates to problem record: twice per day, unless target date defined Target resolution: 2 Business Days

Severity Level P3
Other Calls, General queries, incident not affecting service availability Response Updates to problem record: daily, unless target date defined 5 Target resolution 7 Business Days.

Service and Change Requests

Our service follows our request fulfilment processes to manage service requests and change requests. Where necessary we will raise a call on the service cloud provider, all contact will be managed by us.

Account Management

We assign a named account manager to manage your requirements; this individual will assign suitable resources to assist with Technology advice, engineering and support services. Your account manager will also arrange quarterly service reviews (typically delivered via video call).

Associated Services

Cloud Economics Consultancy

We provide consultancy to assist you in finding the right cloud solutions(s) whether that be AWS, Microsoft, or Private Cloud. We can assist with cloud road mapping and strategy, and economically review any existing cloud presence you may have, to ensure that you are getting the best value for the workloads you have deployed.

Cloud Design Services

Cloud design and deployment – we can design (HLD & LLD) all of the elements of your cloud requirements in any of the Cloud technologies supported by the service.

Cloud migration services are also available as part of our service, as either an assisted or professional services engagement.

Cloud Backup and BCDR

We can provide a born in the cloud solution for Cloud (and non-cloud) backup and recovery if required. The solution is Cyber aware and allows for recovery to any target, on premises, public or private cloud. The solution is a next generation solution that enables restore to different targets than the backup, assisting you with both site recovery and business continuity and disaster recovery scenarios.