FLARE
Information Management

How can Digital Transformation help me handle information overload?

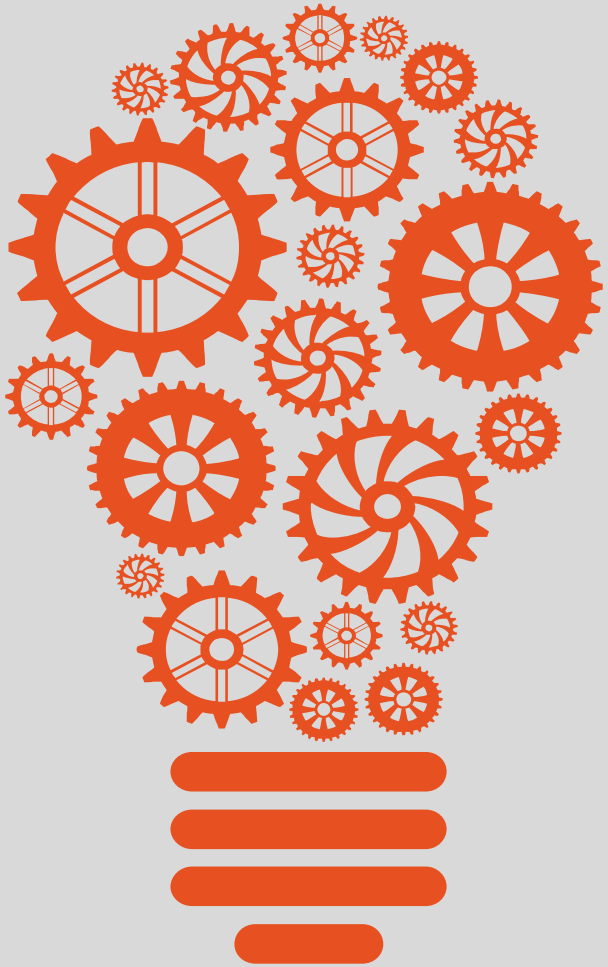
(4 Petabytes incoming!)

Dave Camden, Director

2nd October, 2019

Final V1.0

Case study – Supermajor acquisition



Digital Transformation

Initiatives must consider how vast stores of legacy information can integrated into the future digital environment.

Case Study

This case study will look at a company acquisition and shows how a large incoming data set was handled.

Applicability

The methodologies and approaches could be applied to any digital transformation project.

The challenge

Vast amounts of structured and unstructured data.

Desire to do clever stuff with it.

How can it be done?

Holistic Information Management

- Consulting
- Software
- Services
- Oil & Gas Experience

FLARE

Evergreen solutions

- Operate more efficiently
- Make better decisions
- Reduce risk
- Augment existing solutions

Flare was formed in 1998 and has been evolving and applying its holistic information management approach since then

Clients include supermajors, independents, governments, standards bodies and service companies

The project: 4 Petabytes incoming!

Business Context

- Company acquisition: A multinational **“Supermajor”** oil and gas company acquired a large oil and gas company **“OilCo”** to strengthen its portfolio

Objectives

- Merge, reconcile and re-distribute information assets
- Deletion of non-entitled datasets
- Deriving value from data

Challenges

- Limited timeframe
- OilCo had over 20 years of legacy data
  - 300m+ unstructured files (4 Pb) in two locations (3Pb and 1Pb split)
- Non-connected infrastructures

Goals and Requirements

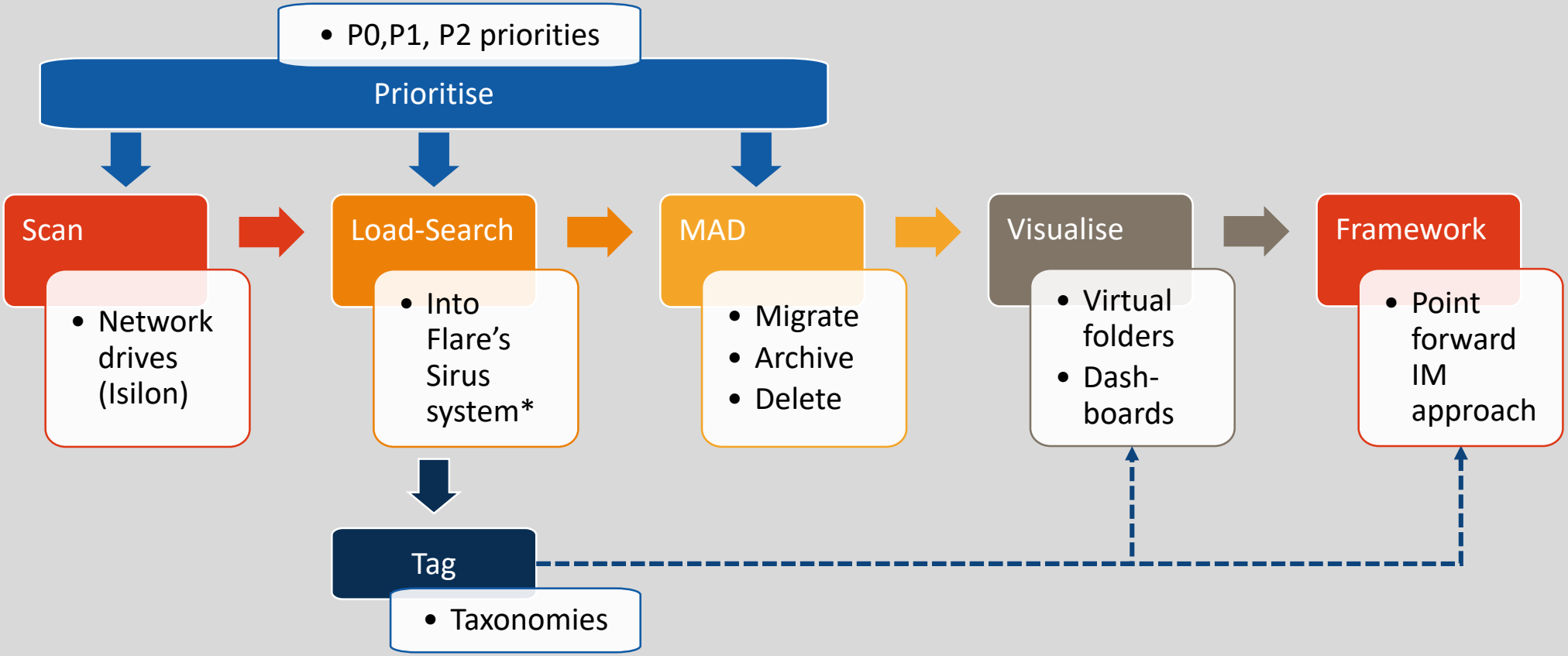
Goals

- Organise and make available high value information from OilCo
- Maintain business continuity

Requirements

- Manage project
- Communicate with all stakeholders
- Create searchable, integrated index of 300 million + items
- Find, prioritise and label ‘information packages’ for Migration, Archive or Deletion (MAD approach)
- Transport Migration and Archive packages to destinations
- Build new environment for incoming information

Approach



** Sirus is Flare's graph based Information Management solution for Oil and Gas*

Disk Scanning

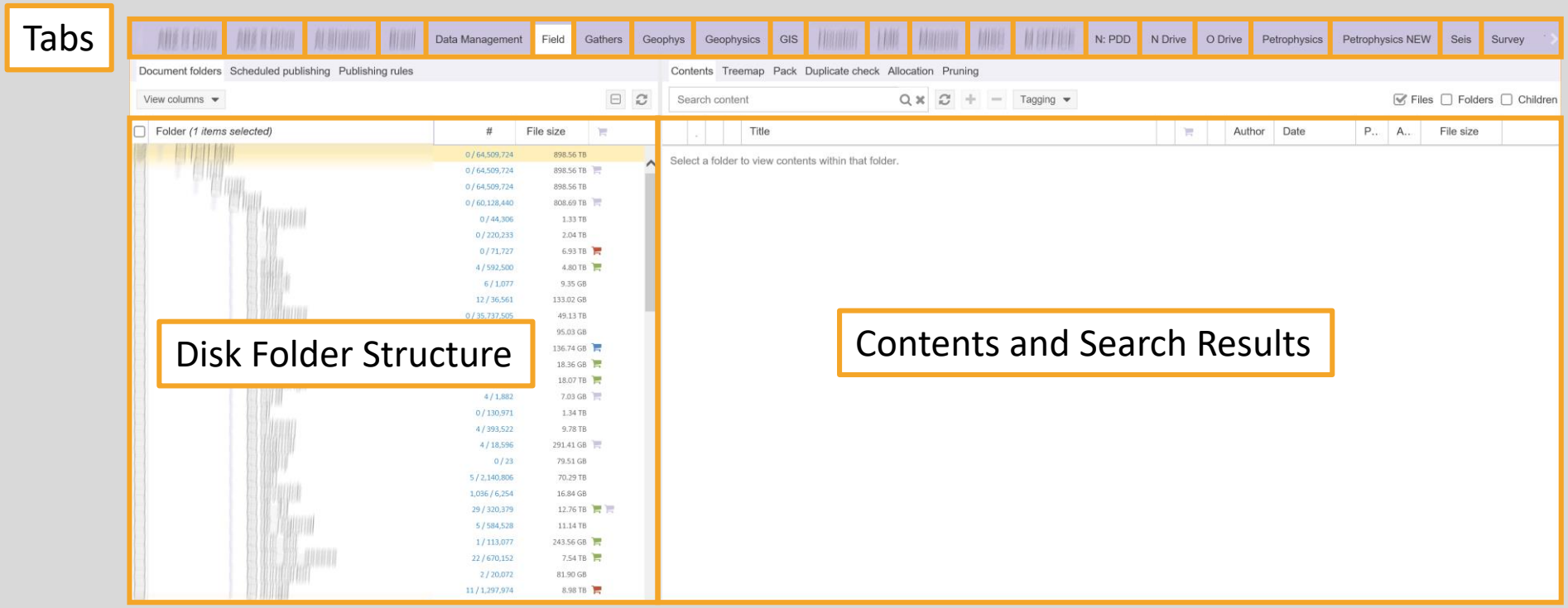
Infrastructure, accessibility and sheer scale meant a pragmatic approach to disk scanning was required

- Disk areas prioritised in 3 groups: P0, P1 and P2
- Disks scanned for main file/folder attributes from the Isilon system
- No checksums were calculated
- No content reading / scraping
- Scan files run through Flare's QC process to detect and fix issues

Scan Loading

The quality-controlled scans were loaded into a graph database system that allowed the original disk structures to be visualised and analysed. URLs link to the disk files.

The 300m+ items were organised based on client requirements into different disk areas on a tabbed interface. Each community can gain access to what they need. Tabs visible based on user.



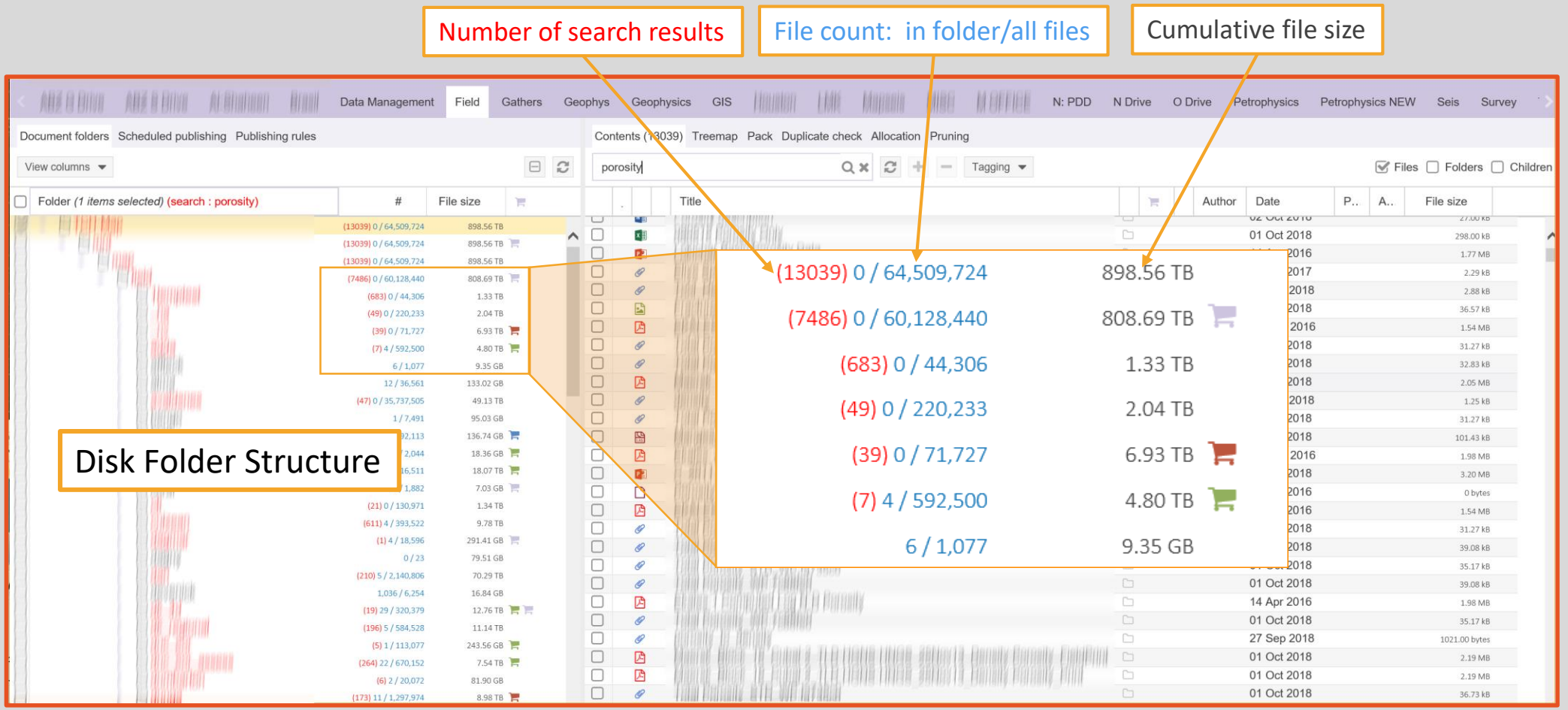
Searching

- The 300m+ items were organised into different areas on a tabbed interface

- Initially searching used string matches on path/file names
  - The key benefits are speed, document counts, roll-ups and folder search/distribution
  - ***The results surprised the client – able to search the entire data set in seconds***

- Search display:
  - LISTING: the familiar ‘results listing’ with links to which folders files are in. Fine for small numbers of results.
  - DISTRIBUTION: where results number 1,000’s – 100,000’s, an innovative and highly performant ‘display search results in folders’ approach was designed by Flare. This visually highlighted which disk areas contained the bulk of the search results, facilitating the partitioning of the disk content.

Flare Folder Search Results

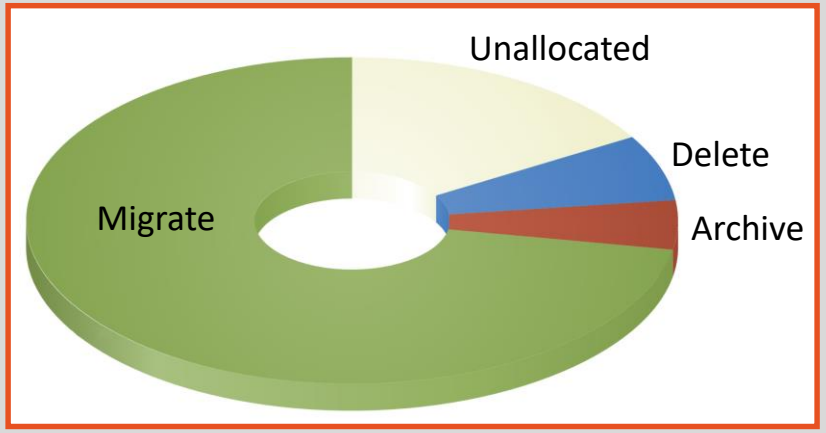


MAD Partitioning

- Used a drag/drop into ‘basket’ system to collate ‘folder hierarchy lists’ (areas of the disk structure)
- Basket types are ‘Migrate’ ( ), ‘Archive’ ( ) and ‘Delete’ ( )
- The baskets are permanent, and record
  - what information has been selected
  - what process is to be carried out

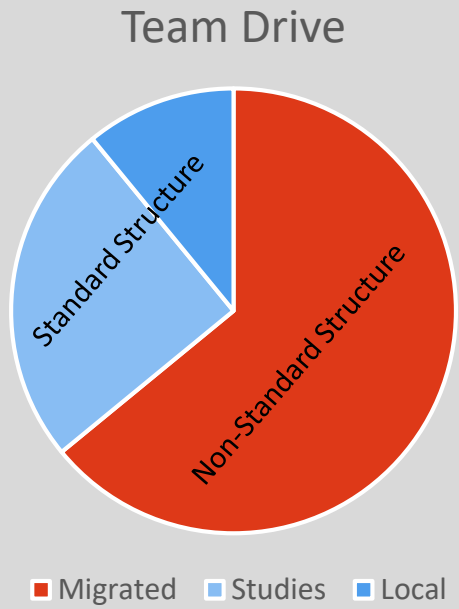
(13039) 0 / 64,509,724	898.56 TB
(7486) 0 / 60,128,440	808.69 TB
(683) 0 / 44,306	1.33 TB
(49) 0 / 220,233	2.04 TB
(39) 0 / 71,727	6.93 TB
(7) 4 / 592,500	4.80 TB
6 / 1,077	9.35 GB

- The overall progress of each disk area is monitored to measure project progress
- The resultant lists are passed on to the IT service group to carry out the required action



Reception of Incoming Data

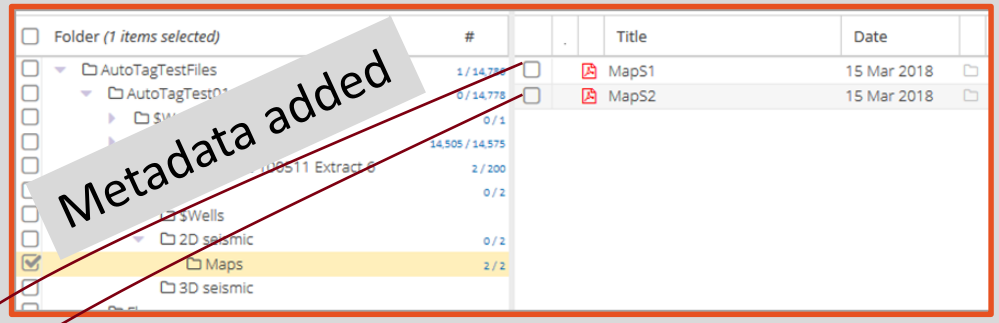
- Supermajor regional office received their Migrated data and loaded to new Team Drives. Windows for working files, Unix for application data
- Data for Migration to other Supermajor business regions was channelled through a centralised data handling group
- Regional office populated new ‘team-drive’ from various sources
  - Migrated data from the 4Pb drives – in original structure
  - Study groups – in ‘company standard’ structure
  - In-situ, on-going work – also in ‘company standard’ structure
  - Approx. 40m files



Re-structuring with Virtual Folders

- Resulting folder structure a mix of ‘company standard’ plus what came from Migration
- Decision made to keep ‘original’ folder structures on disk …
- **.. and create high-level additions in Flare Sirus using virtual folders**

Sirus folders representing Isilon structure

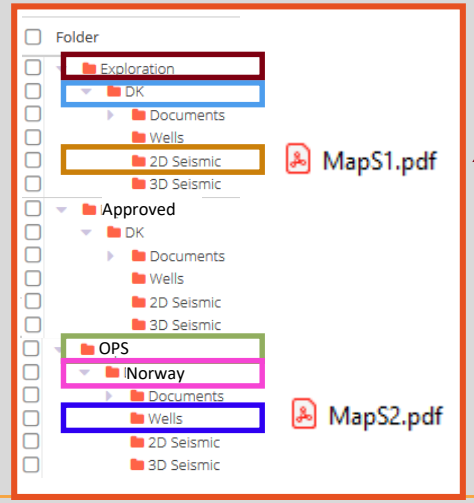


MapS1 = Denmark ‘2D seismic’ map EXP-G

MapS2 = Norway ‘well location’ map OPS

- A virtual folder shows content which is tagged with specific, standardised metadata
- ‘Target’ folders were identified and tagged using a mix of methods :
  - Manually
  - Automatically

Sirus Virtual folders



Appears in

Results

Outcome so far

- Users have access to multiple data sets with one search
- Successful integration with minimal disruption
- P0 and P1 migration completed over a 6-month period
- P2 is part-complete and ongoing
- Audit record of what happened to 300m files
- New Regional Team Drive
  - has stayed operational throughout
  - has been ‘live scanned’ and is continuously monitored to maintain an up-to-date index
- Have built a foundation for moving forward

In Progress/Next Steps

- Index and tag team drives using standardised terms from a set of taxonomies (actually an ontology)
  - ‘Asset’ taxonomies of wells, fields, licences, countries etc.
  - ‘Context’ taxonomies based on deliverables (work products) plus relationships to Topics, Disciplines etc.
  - All taxonomies hierarchical, inter-related and include synonyms
- Tagging
  - Manual
  - Business Rules
  - Inheritance – see below
  - Automatic – joint analytics / ML program to improve accuracy and capabilities
- Continue drive monitoring to capture any changes on the file system
- Hide ‘uninteresting’ areas
- Modelling key processes
  - Identify deliverables by process step
  - Simplify tagging using drag/drop and inheritance (leveraging Flare’s comprehensive taxonomy)
  - Simultaneously track progress of project vs deliverables

Learnings (i)

- Project controls
  - Develop a bespoke way to keep a track of progress accounting for large data volumes
  - Partition the data and prioritise based on business needs – involve end-users

- Be selective to avoid information overload
  - Don’t handle everything
  - Work at folder-, not file-level where appropriate
  - ‘Hide’ low-level application files

- Big network-based data sets (100m files / Petabytes)
  - Require a fast search system - native file systems often too slow/inflexible to be practical
  - Indexing files and moving to alternative storage takes a LOT of time

Learnings (ii)

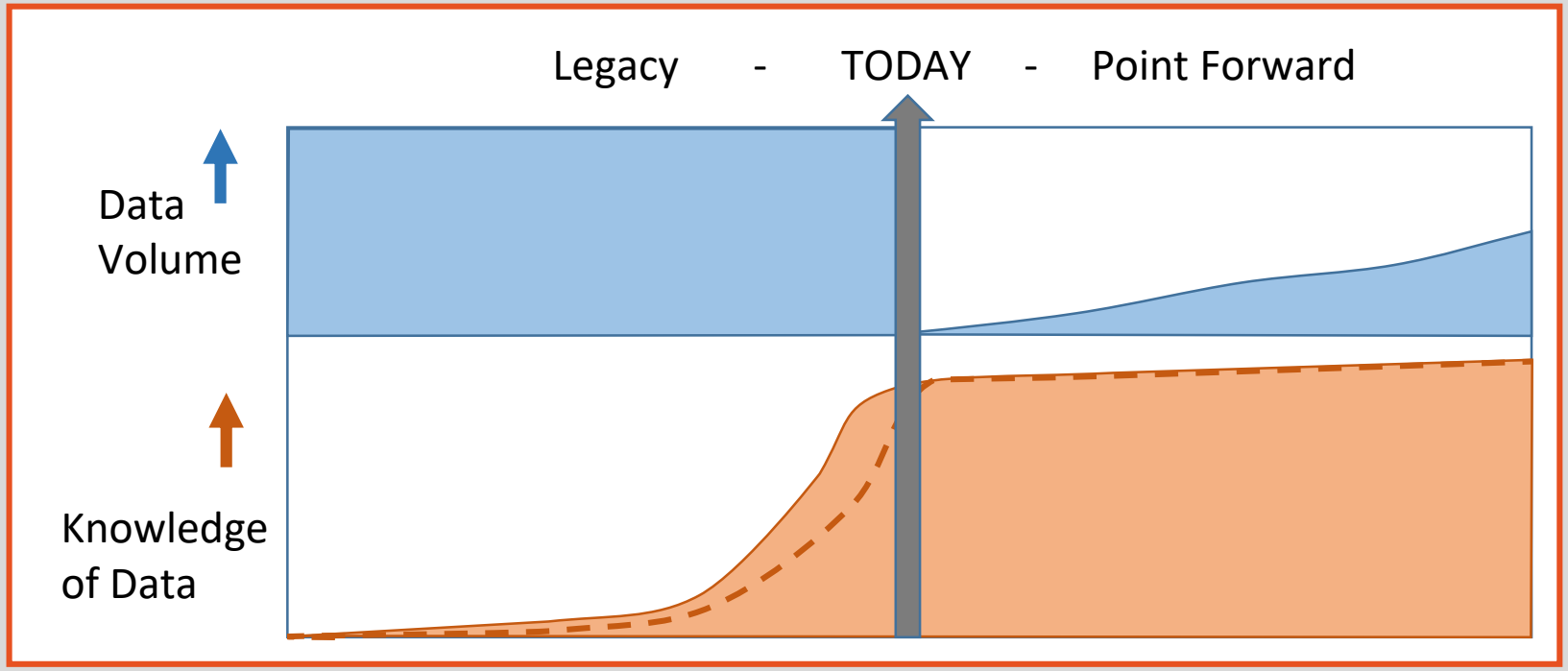
- Keep it simple
  - Milliseconds count! Use the fastest, simplest methods initially
  - Use more complex methods progressively
  - Don’t aim for 100% - reasonable endeavours then improve incrementally

- Folders
  - Useful for project-based working, but need augmenting with metadata and ‘smart’ searching. Folders are less useful for enterprise-wide storage.

- Deletion
  - Data past their retention period or no longer useful should be deleted, but must know what you have in order to do this – good metadata is essential, especially point-forward
  - Keep a record of what has been deleted

Legacy and Point Forward Information

Legacy

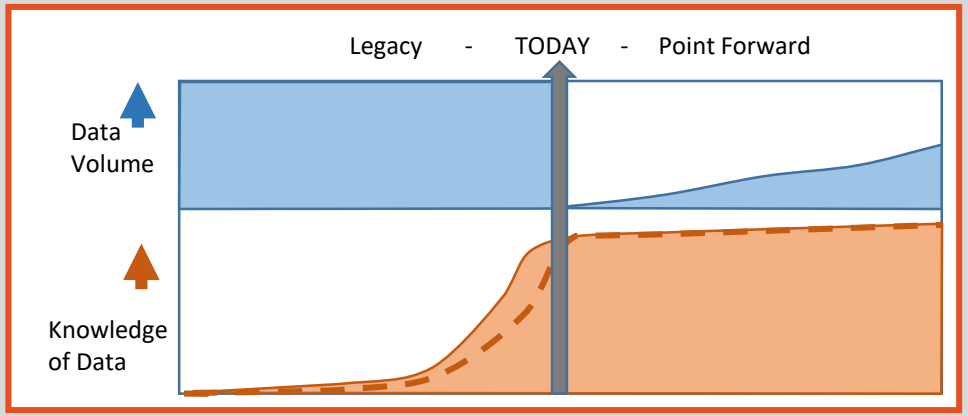


- Work so far has been on the legacy information: this system index will stay in place to provide an historical searching capability and audit record of what was done
- Other legacy/current information sources are being tagged to provide an integrated search across multiple information systems
- Large, often poorly structured data sets with little metadata
- Original creators are gone! Automate as much as possible

Legacy and Point Forward Information

Point Forward

- Creators are in place - opportunities for knowledge capture
- Opportunity to create better-structured, higher-value data sets
- Don’t rely on folder structures alone – also use metadata tags and virtual folders to create alternative views, dashboards and support searching
- Understand concept of ‘information objects’ that are useful to the business
- Add standard metadata to support management and searching



Benefits and Enablers

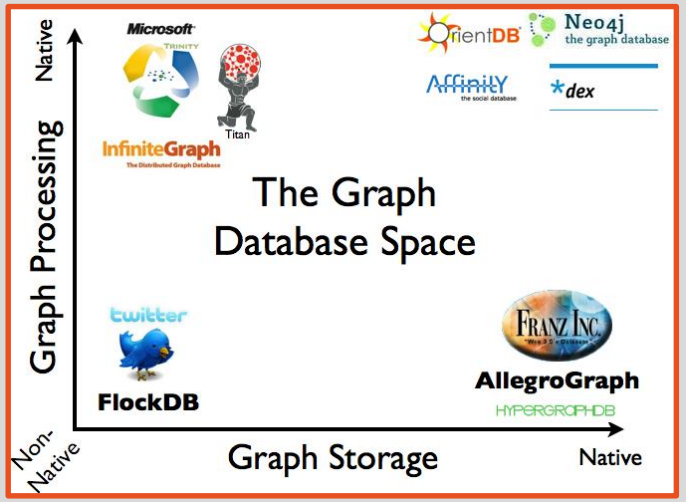
Benefits

- Unified corporate information assets
- Teams' information needs effectively addressed
- Rapid search capabilities
- Custom views
  - Virtual folders
  - Information objects
  - Search 'in folders' (distribution)
  - Hiding uninteresting
- Foundation for the future!
  - Integration, support analytics

- This project was dealing with traditional company-based infrastructure, but the approach is relevant to Cloud-based environments
  - Can you easily move 4Pb of content and its metadata from AWS to Google Cloud?

Enablers

- Graph Database



- Metadata tags
- Oil and Gas taxonomies
  - Asset
  - Context
  - Other taxonomies
- Holistic IM approach

FLARE Taxonomy

Thank you for listening. Any questions?

Flare Solutions Limited
3, Acorn Business Centre, Northarbour Road,
Cosham, Portsmouth, PO6 3TH, UK

Tel: +44 203 397 7766
Fax: +44 870 460 2543

enquiries@flare-solutions.com
www.flare-solutions.com

Dave Camden
Director

d.camden@flare-solutions.com
+44 7703 234 891
+44 1892 785 007