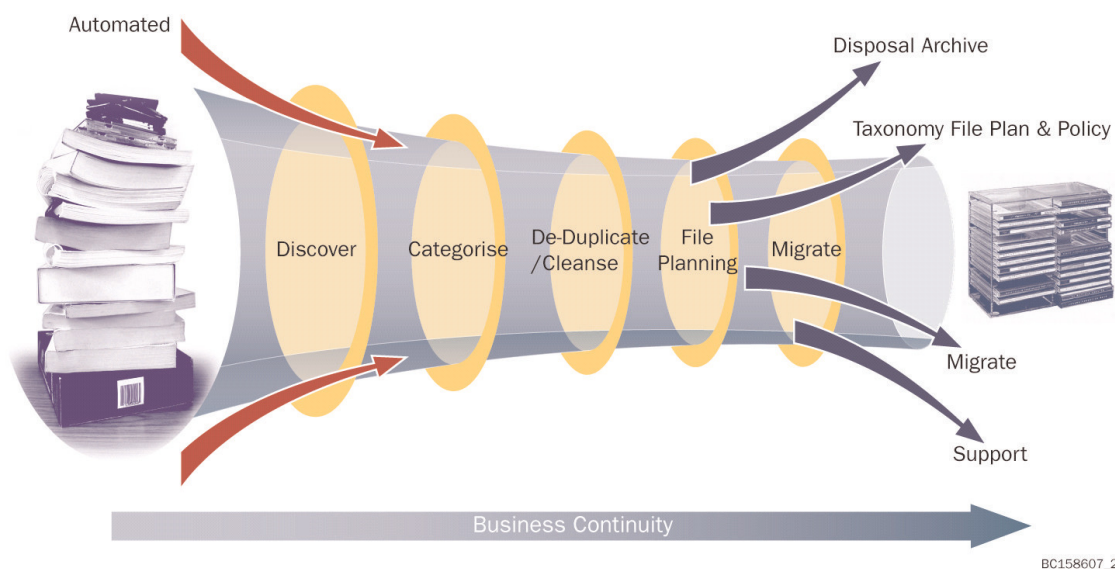


Scope of Pilot Data Discovery and Content Classification Service



Clarity uses the Active Navigation information discovery tool to support the categorisation of information prior to migration to a formal, managed records store. During this service, we perform a linguistic analysis of existing stored content to identify the main areas of business that the content addresses and all duplicate, near-duplicate and other redundant data files (experience has shown that typical levels of unnecessary replication vary from 20% to above 60%). An initial pilot analysis is used to span up to 100,000 files that are stored in a variety of formats, such as PDF, DOC, JPG, MP3, etc.

We also use “themes”, which are extracted during the audit, to provide business content-driven guidance for relevant subject matter experts, whilst defining classification structures under which to store the newly-cleansed content (e.g. to support implementation of a Business Classification Scheme).



The data discovery and content classification service comprises the following key stages:

Discovery

Clarity performs an initial analysis of the selected data holdings in order to identify the types of files that exist, the size and volume of file types and potentially redundant and irrelevant content. As part of this process, Clarity creates reports that are provided for review by subject matter experts and content owners, to determine what data may be archived.

Full Content Analysis

After the initial discovery, and cleansing, process a deeper analysis of remaining content is run. This process will:

- Extract keywords and phrases from the full text of the content (no requirement for prior knowledge or dictionaries required)
- Generate intelligent summaries from the full-text of the content (those sentences that best describe the context of the document)
- Calculate duplication statistics and reports for further review
- Assign meta-data values based on any required attributes.

Duplication Cleansing

Clarity uses the Active Navigation Duplication Report Editor to review the results of the duplication analysis, in collaboration with nominated individuals from the target business area/s. This application uses Windows logon credentials to ensure that users can only ever see data that the network permissions entitle them to see.

As part of this exercise, Clarity demonstrates how users can initially recommend what to do with the results and an administrator can then review and accept recommendations using the Duplication Cleanser. At this point sample files that are deemed surplus to requirements may be moved to a holding area automatically.

File Plan Development

This stage of the process enables an organisation to review the output from the full content analysis phase and use this to create a new, simplified folder structure, which reflects the content as well as the organisational requirements.

The following tasks are performed during this stage:

- Clarity creates a skeleton File Plan structure, based on the findings from the full content analysis.
- A review meeting or workshop is held with nominated staff from the business area that is being processed in order to discuss the results from the analysis, review the proposed new structure, propose modifications and agree a final “approved” version.
- Once the final folder structure has been agreed Clarity creates the new folder structure on disk and demonstrates the migration of sample content to the new structure using the Active Navigation Content Migrator. This process will NOT apply security permissions automatically: enabling an organisation to rationalise their security model and impose some order on what may previously have been chaos.

The Content Migrator may subsequently be used by suitably authorised personnel to move content quickly from an old structure to a new structure.

Users can bulk move content or review files on a case by case basis. Windows login credentials are used to ensure that personnel can only see data that their network permissions entitle them to see.

It is important that users are involved in the migration process, since a wholly automated approach to migration can cause difficulties if users are unable to find relevant information.

Reporting

Following completion of the data discovery, analysis and collaborative file planning activities, Clarity produces a report that describes the findings and recommendations derived from the exercise.

Installation Requirements

Clarity shall need to install Active Navigation software on a server, or a networked desktop PC and we recommend a minimum hardware specification of 1 GHz processor, 2 GB RAM, 10 GB free disk space.

Clarity provides SQL Express as the database repository for use during the pilot.

Price

The total fee for provision of a pilot Data Discovery and Content Classification Service is £5,000.

Clarity Information Solutions

Stewart House
South Grove
Rotherham
South Yorkshire
S60 2AF

Tel: 01709 367028
Fax: 01709 367029
E-mail: sales@clarity-ltd.co.uk
Web: www.clarity-ltd.co.uk

