Central Digital & Data Office

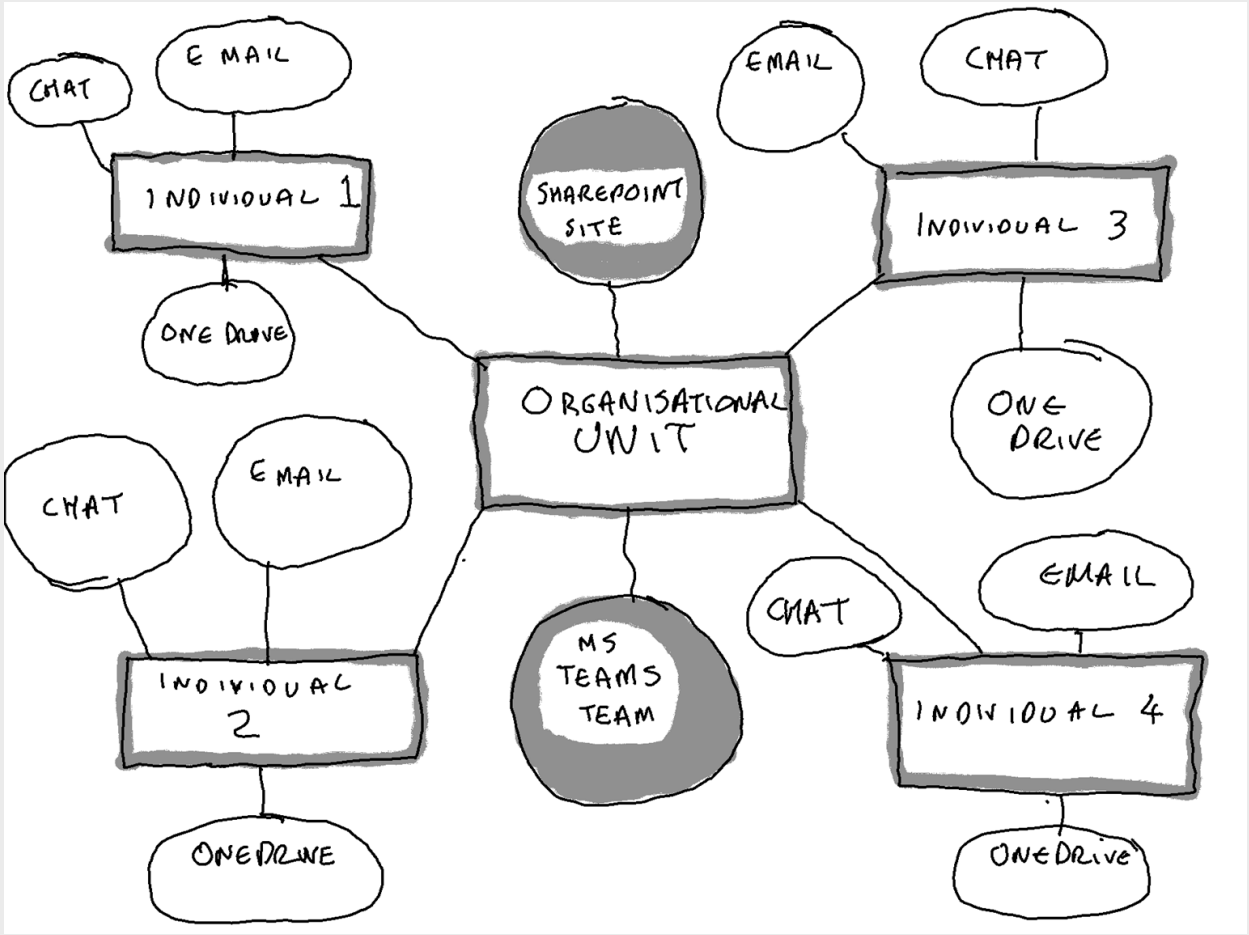
Principals and pathways for the application of AI to records management

James Lappin

Presentation to DLM Forum members' meeting

28 May 2024

- **Aim of this presentation**
- Tasks for AI
- Relevant records management principles
- A pathway
- The use of AI within the pathway


CHAT
E MAIL
INDIVIDUAL 1
ONE DRIVE
SHAREPOINT SITE
EMAIL
CHAT
INDIVIDUAL 3
ONE DRIVE
ORGANISATIONAL UNIT
CHAT
EMAIL
INDIVIDUAL 2
ONEDRIVE
MS TEAMS TEAM
CHAT
EMAIL
INDIVIDUAL 4
ONEDRIVE

Corporate digital systems have structures.

Those structures are used to apply predictable access rules and retention rules to content.

There is a need to identify and test pathways to enable more precise application of access and retention rules.

AI will be a component part of those pathways

This presentation sets out one potential pathway

- Aim of this presentation
- **Tasks for AI**
- Relevant records management principles
- A pathway
- The use of AI within the pathway

It would be beneficial to have an AI capability to group together aggregations (SharePoint sites, shared drive folders, email accounts etc.) that have arisen from an activity or a similar range of business activities).

It would be beneficial to have an AI capability to look at a given aggregation and group together items (documents, messages etc.) within it that have arisen from the same business activity.

It would be beneficial to have an AI capability to distinguish unsolicited, trivial, uncommunicated, personal and social content from business content

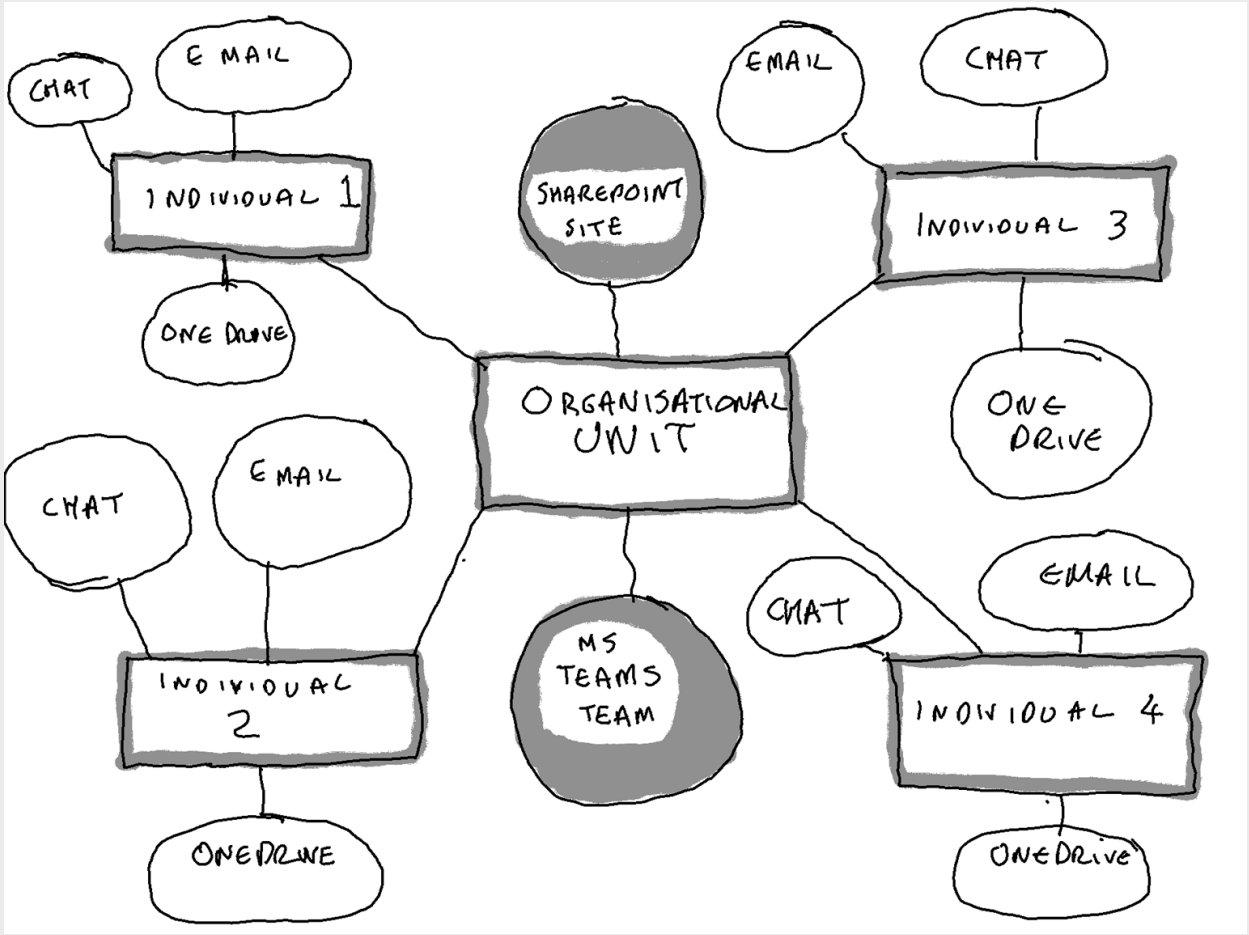
- Aim of this presentation
- Tasks for AI
- **Relevant records management principles**
- A pathway
- The use of AI within the pathway

Records management aims to preserve the context of records as well as the records themselves.

The sites/accounts in which records were created and received form part of the context of the records.

Good record systems enable an end-user to predict what will happen to their content over time.

Good record systems give end-users a broad idea of who is likely to be able to access their content and how long it will be kept.


CHAT
E MAIL
INDIVIDUAL 1
ONE DRIVE
SHAREPOINT SITE
EMAIL
CHAT
INDIVIDUAL 3
ONE DRIVE
ORGANISATIONAL UNIT
CHAT
EMAIL
INDIVIDUAL 2
ONEDRIVE
MS TEAMS TEAM
CHAT
EMAIL
INDIVIDUAL 4
ONEDRIVE

Imagine a situation in which three different teams work separately on a particular policy issue. Each team has their own space/site in which they create and store content.

If AI was later to be used to merge content from the three shared spaces, then this would risk creating a false impression that there had existed a combined space in which those teams had shared content with each other.

- Aim of this presentation
- Tasks for AI
- Relevant records management principles
- **A pathway**
- The use of AI within the pathway

Pick the level of aggregation that will serve as the point of continuity between the structure that end-users worked in, and the improved structure that you will end up with.

It should preferably be the aggregation through which default access rules were applied. Examples include SharePoint sites, folders in shared drives and email accounts.

In the pathway AI can be deployed to:

- Create an overall classification of the aggregations (to group aggregations from similar activities together)
- Create sub-aggregations within those aggregations (like email accounts) that contain records of many different activities

In this pathway AI is not deployed to:

- Move items from one aggregation to another

- Aim of this presentation
- Tasks for AI
- Relevant records management principles
- A pathway
- **The use of AI within the pathway**

Data clustering algorithms could be used to map the proximity of different aggregations (sites, libraries, accounts) to each other.

This is a ‘bottom-up’ way of generate a classification framework that could enable:

- aggregations to be placed in context
- coherent retention decisions to be made across the digital heap

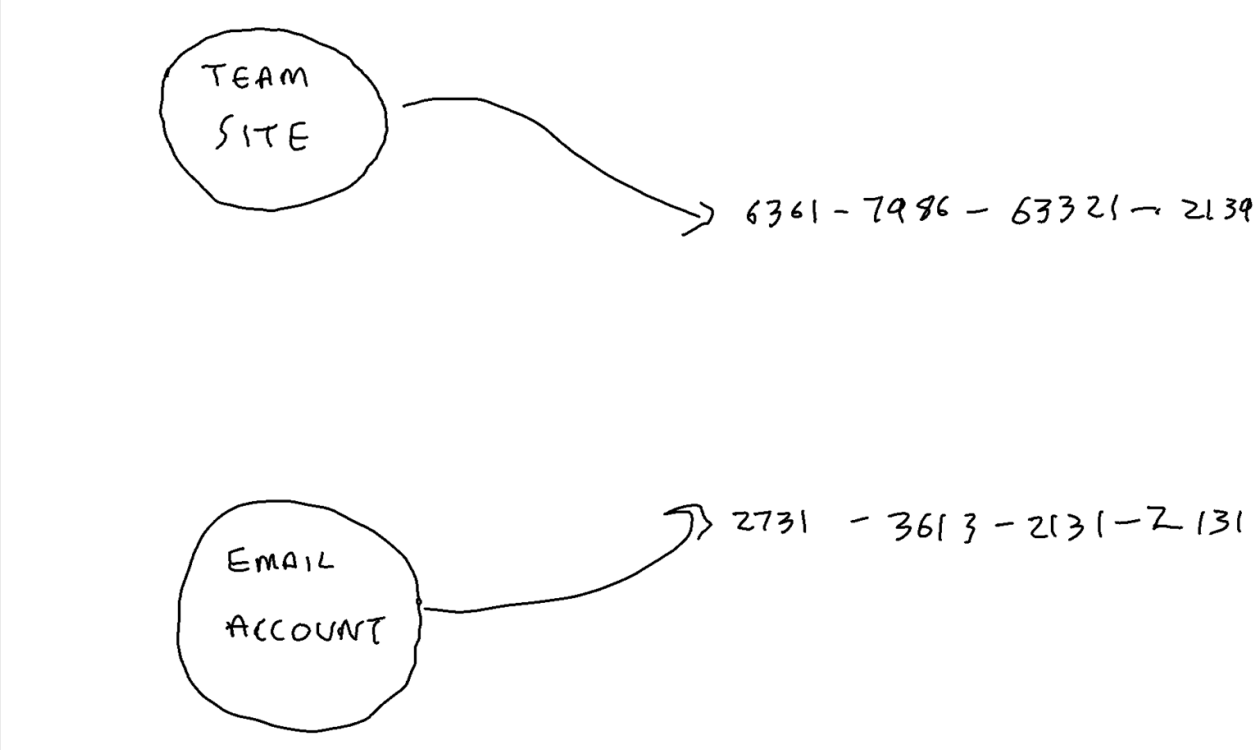
The rationale behind data clustering algorithms is as follows:

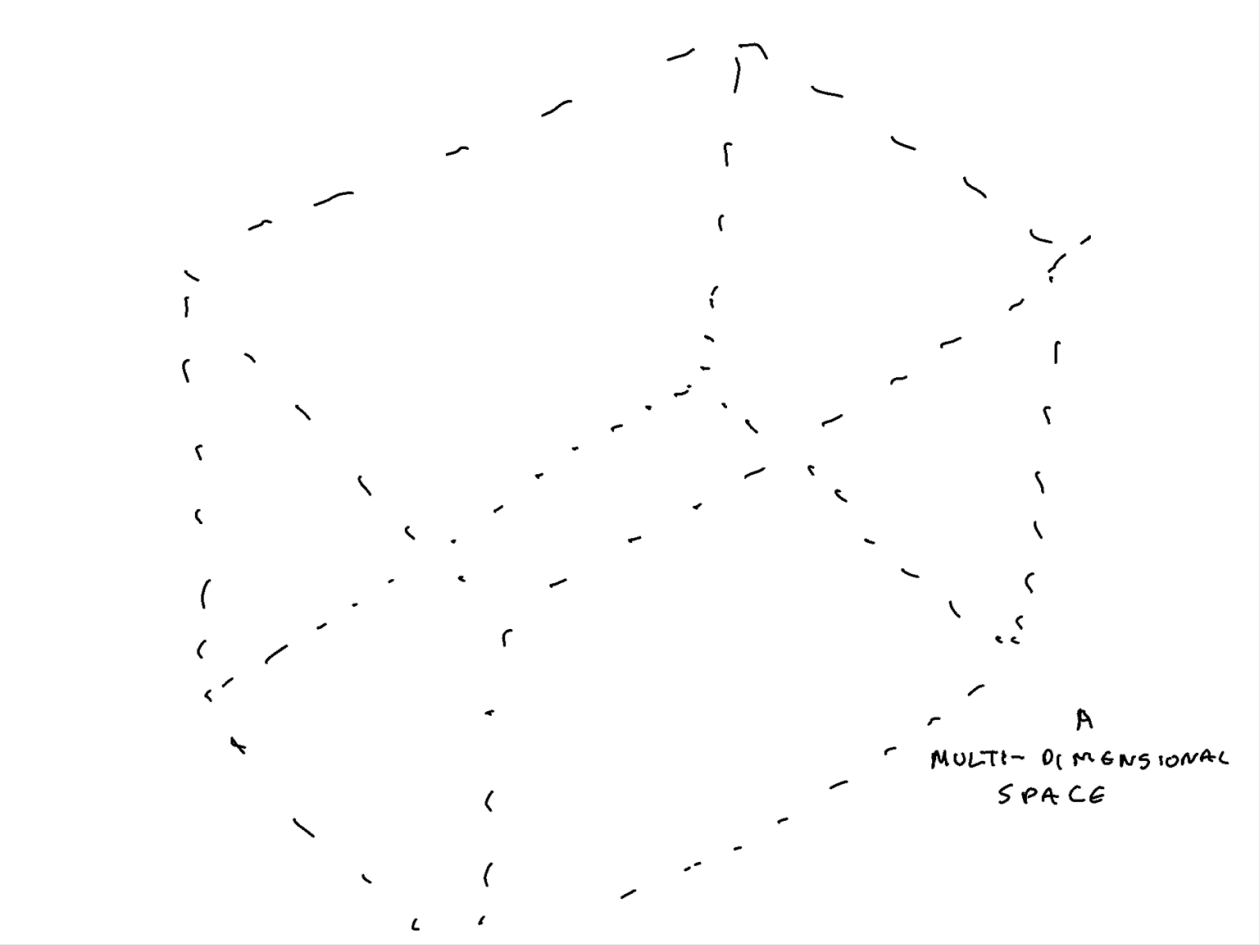
> Given a set of data points, partition them into a set of groups which are as similar as possible.

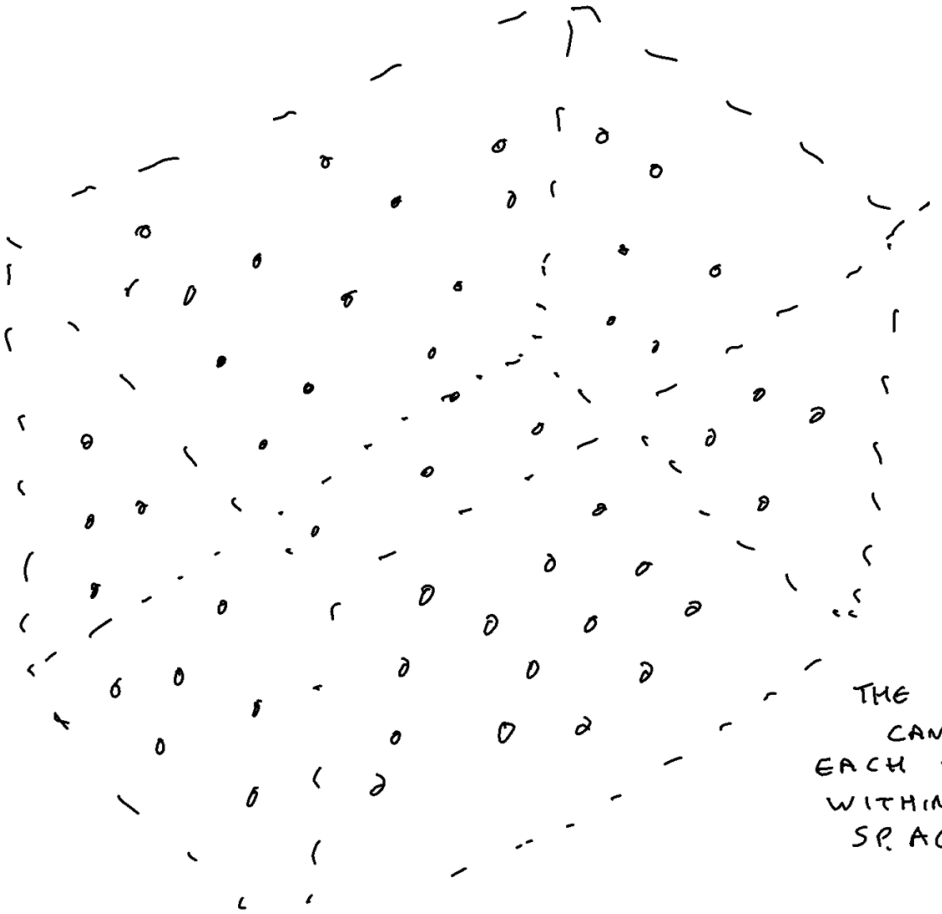
Data clustering algorithms require just as much supervision as any other approach to AI.

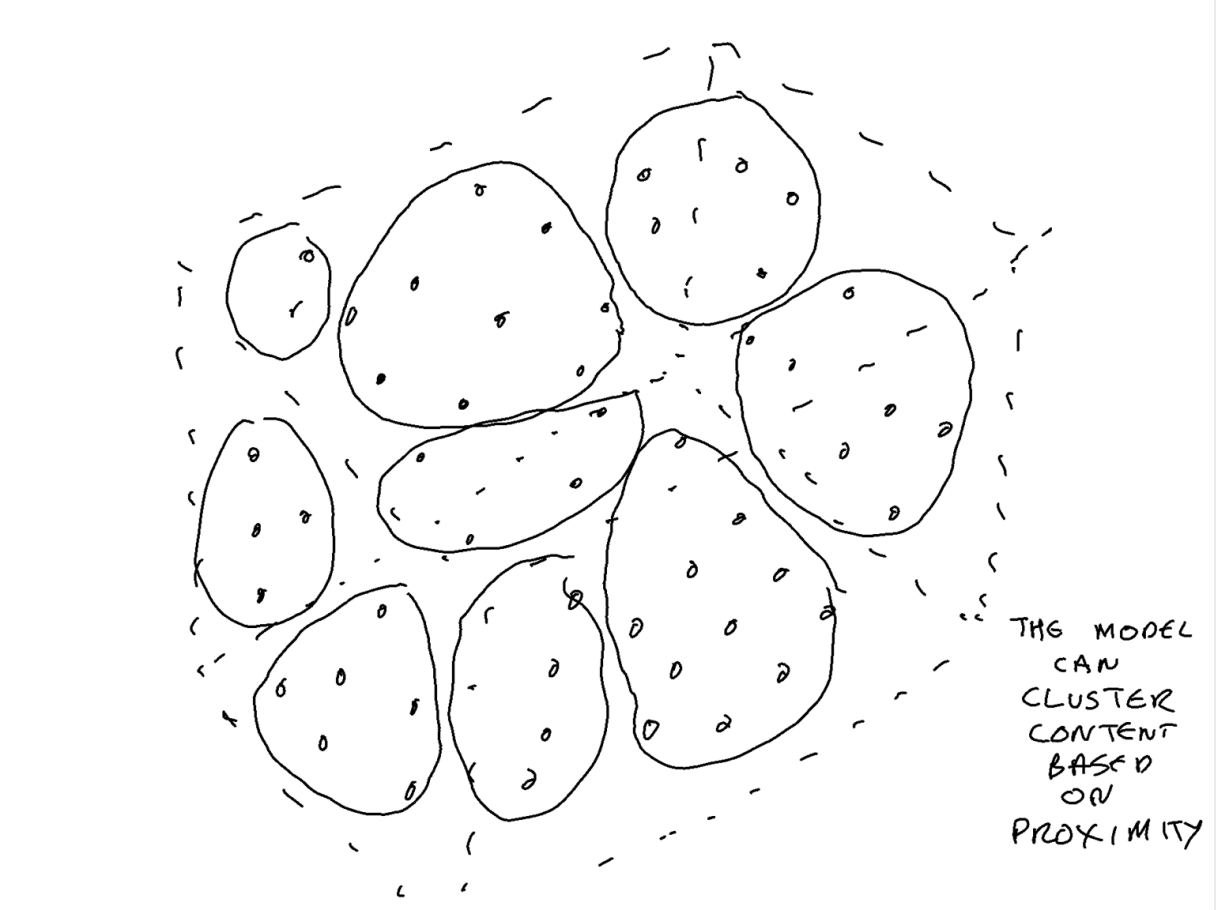
Any use of AI to change the basis on which access or retention rules are applied to records should take place under supervision by records professionals.

Aggregations of documents and messages can be turned into vectors

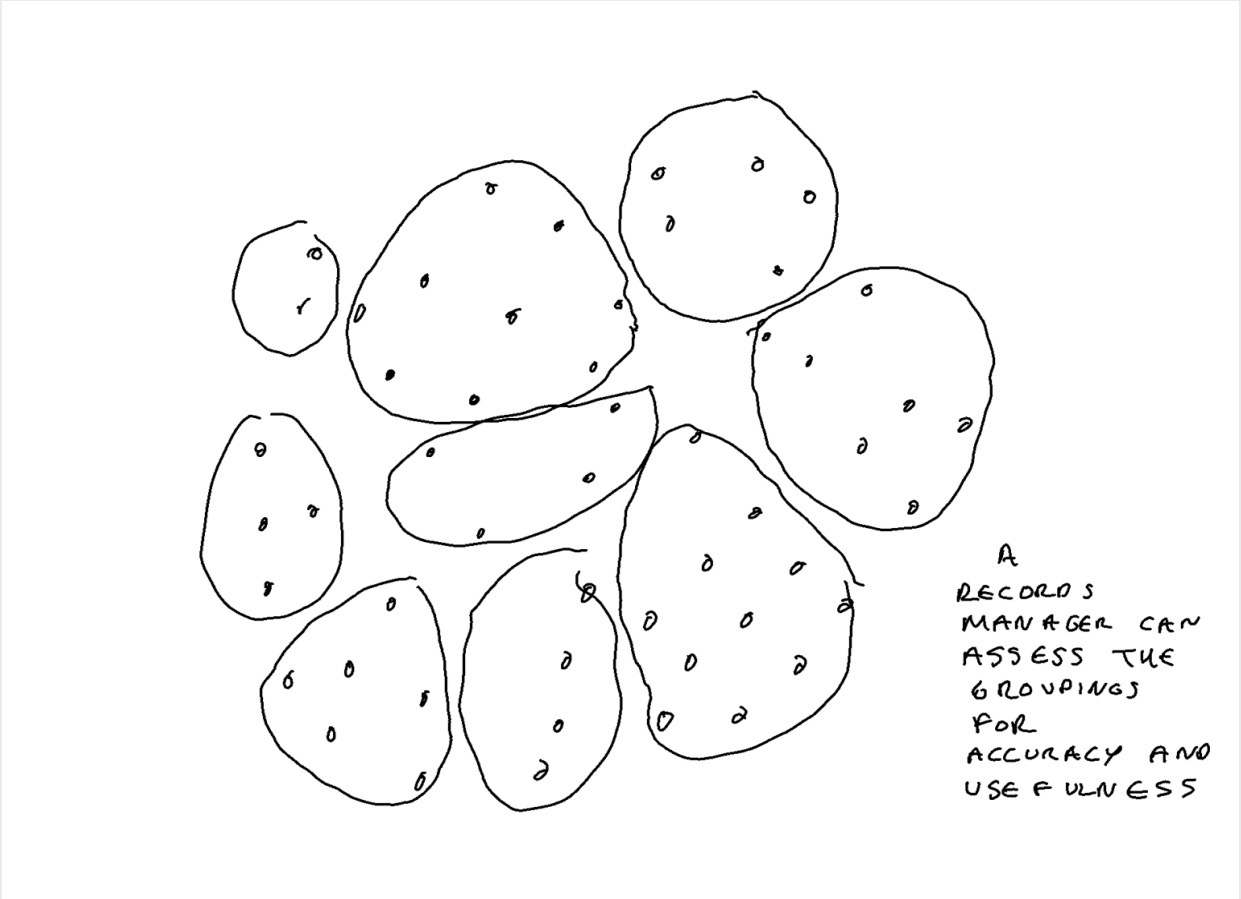



A
MULTI- DIMENSIONAL
SPACE


THE MODEL
CAN PLOT
EACH VECTOR
WITHIN THE
SPACE


THE MODEL CAN CLUSTER CONTENT BASED ON PROXIMITY

The aim is to arrive at groupings that bring together email accounts, team sites and/or teams that had arisen from a similar function or range of activities.



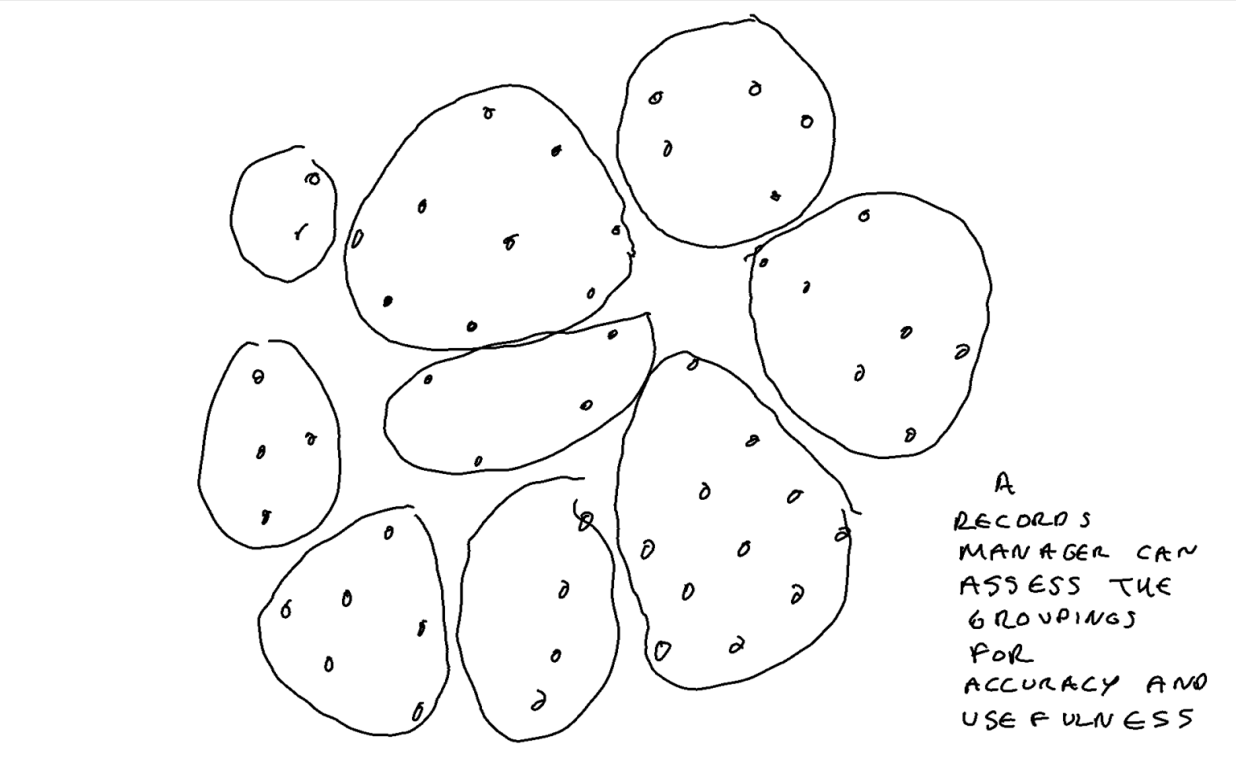
Data clustering algorithms could be used to create new clusters within problematic aggregations, such as email accounts.

DOCUMENTS AND MESSAGES CAN BE TURNED INTO VECTORS

59133 - 39543 - 345371 - 4132

61837 - 4372 - 4376 - 43722

Related correspondence within an email account could be clustered together under a common tag or classifier.



Related correspondence within an email account could be clustered together under a common tag or classifier.

This would support the filtration of trivial and social content from within individual accounts.

It would also offer a pathway for individuals to grant their successor-in-post access to business content within their individual account.

Central Digital & Data Office | Cabinet Office

Central Digital & Data Office
For further information about contact
james.lappin@digital.cabinet-office.gov.uk

CDDO