



NHS ENGLAND

DATA CATALOGUE REVIEW

3555

Industries, Location of - N
Industry - Social Aspects - B

3566

Information The
Inga

3567

Ingb
Inglir

3556

Industry - Social Aspects - C
Industry and State
(by place) G

3568

Inglis
Inheritance and

3557

Industry and State
(by place) H
Indyh

Inheritance an
Injections, R

Mike Thompson, mySociety

6th February 2015, Version 1

INDEX

Introduction	3
Methodology	4
Users	5
User Journeys	8
Usability Issues	9
Areas of Improvement	17

INTRODUCTION

INTRODUCTION

Mike Thompson from mySociety has been commissioned by NHS England to conduct a review of the nascent NHS England Data Catalogue. This review is intended to find areas where the design and functionality can be improved based on direct feedback from potential users.

Mike has been conducting interviews with Data Catalogue users to understand more about how they work, how the Data Catalogue might help them in what they do and to get their ideas for how it could be improved, partly through usability testing of the interface.

This document describes the findings from these interviews and identifies potential areas for improvement and further exploration in the remainder of the project.

METHODOLOGY

APPROACH

I have conducted two days of interviews with Data Catalogue users, on the 14th and 15th January 2015.

In these interviews, I asked the participants questions about their work and how they use data to create reports, asked them what they knew of the Data Catalogue and probed what they liked and disliked about it.

I also asked participants to use the catalogue to complete two or three tasks in order to identify potential usability issues with its user interface.

USER INTERVIEW PARTICIPANTS

The eight interviews I conducted were with people in different roles relating to the catalogue. These are shown below. The users that are of most interest to this analysis are shown in a darker colour.

Of particular interest are the analysts who are the biggest consumers of data and the people who carry out NHS England’s routine reporting function.

Startup	CSU Analyst	Analyst	Analyst
Operations	Operations	Analyst	Analyst

USERS

OVERVIEW	USER TYPES		
<p>In the interviews, I spoke to three main types of user that are useful for further analysis.</p> <p>The first of these is the analyst—someone who spends almost all their time finding and using data to compile reports; the second the ops person—someone trying to use data to answer questions about how effectively something is working; and the third the startup—someone interested in using data to create new stories, products or services.</p> <p>This page gives a brief overview of each. The Analyst is described in more detail on the following page.</p>	<div><div>The Analyst</div><div>Weighting: 50%</div><div><p>I need to find the most up to date version of the data I use in my reports</p><p>I hope it will speed up the boring stuff and let me concentrate on the more interesting work</p></div></div>	<div><div>The Ops Person</div><div>Weighting: 30%</div><div><p>I need to find the information that will help me evaluate whether an intervention was successful</p><p>I hope it will act as a central source of data with good metadata and provenance</p></div></div>	<div><div>The Startup</div><div>Weighting: 20%</div><div><p>I need to explore data to find stories can turn into new articles or products</p><p>I hope that having a central source will mean I don't have to try to understand the bureaucracy</p></div></div>

USERS: THE ANALYST

GOAL

To find the most up to date version of the data they use in their reports.

BACKGROUND

The analyst produces reports, many routine, to help operational staff, managers and other NHS organisations understand performance, quality and make decisions.

CHALLENGES

1. I already know where to find the data I use regularly
2. It's hard to tell how good the catalogue is without much data
3. If I can't find what I'm looking, or if I experience issues with the system, I'll go back to what I know

NEEDS

- 1 I need to know coverage and release dates of a dataset to know whether it's really the one I need
- 2 I need to quickly see what a dataset looks like to know if it's what I'm expecting
- 3 I need guidance when submitting a dataset to know what I should enter

FRUSTRATIONS

1. I can't consistently see coverage dates or release dates on search results
2. I can't filter search results by date
3. I'm not sure what sort of thing to enter and how to format title and description when adding a dataset
4. I don't understand why it asks me for name and description again when uploading a file
5. I'll just go to the horse's mouth if downloads don't work properly

USERS: SUMMARY

SUMMARY

The analyst is the most important user in the short to medium term for the data catalogue.

As well as the needs identified for the analyst, there were other needs identified in the research, specifically:

- needing to know **how a dataset is broken down geographically** in order to know whether it will satisfy the analysis the user wants to perform
- needing to **contact the owner** of a dataset in order to ask questions about, for example, provenance, structure or potential mistakes
- needing to **link from a dataset to where it is used** in a report or a dashboard to help people find more detailed analysis and interpretation.

From a project perspective, some of the most interesting findings of this analysis are the challenges that the data catalogue will face around acceptance, specifically:

- being better than the other systems providing the same data or being sufficiently strongly mandated to get around this
- ensuring that the catalogue is sufficiently useful for a group of users before marketing it to them
- making a good first impression: users quickly revert to established ways of working if something doesn't impress.

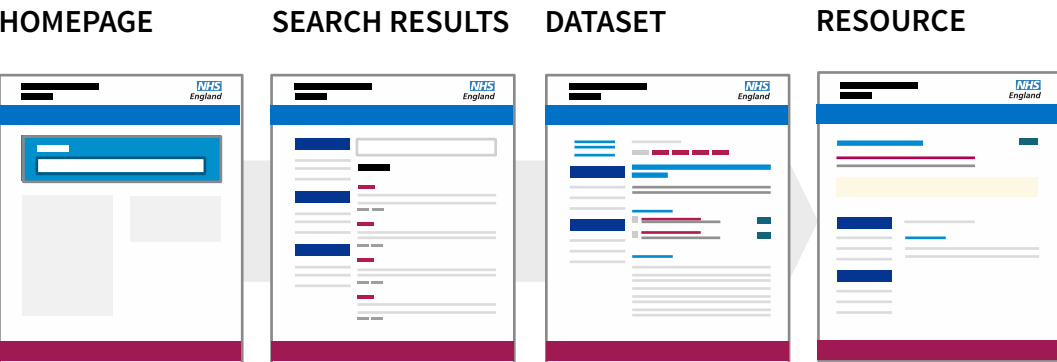
USER JOURNEYS

OVERVIEW

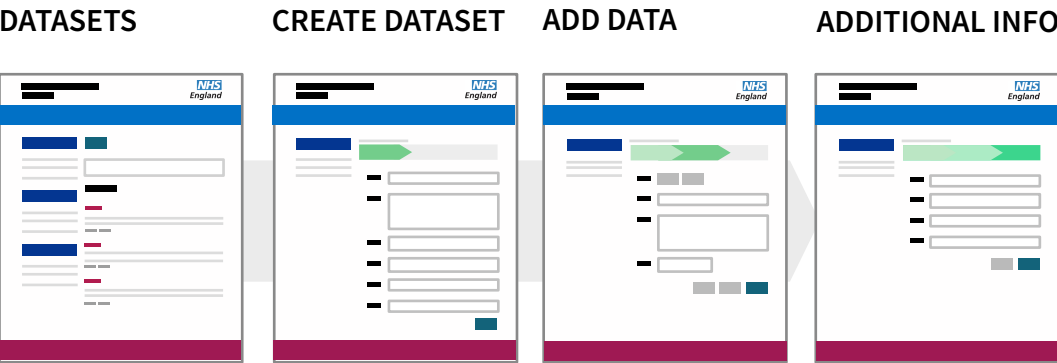
The two diagrams to the right show the visual structure of the two main journeys through the catalogue that were tested.

These illustrations are used on the subsequent pages about usability issues, annotated with details of the problems that users had with them.

FINDING A DATASET AND DOWNLOADING DATA



ADDING A DATASET



USABILITY ISSUES

OVERVIEW

The second half of the interviews focused on evaluating the user interface of the catalogue by asking participants to use it to complete certain tasks.

Users were asked to search for and download a dataset relevant to them and then to add a new dataset.

This section of the report details the problems that users encountered. This page highlights the most serious problems and subsequent pages break down individual problems in more detail.

PROBLEM SUMMARY

- 1

Coverage and release dates are crucial in helping analysts and others find the right datasets but are missing from search filters and from search results pages.
- 2

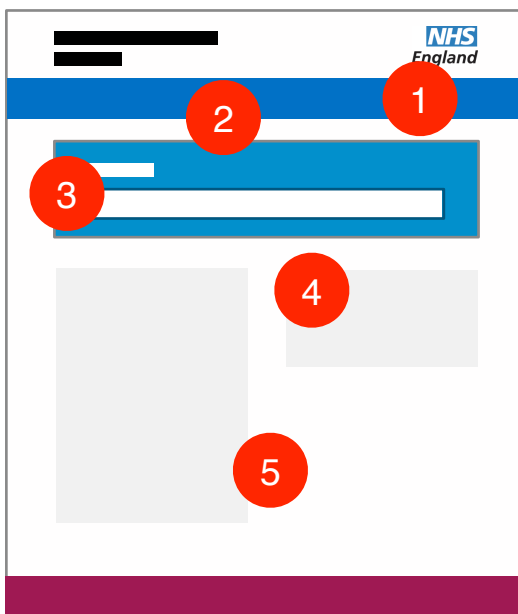
Almost all users tested had problems understanding how to fill in the form to add a dataset, needing prompting and explanation to complete the task.
- 3

Download and view actions on files within a dataset are hidden behind an explore button and change based on file type, unnecessarily confusing users.
- 4

The visual hierarchy and layout of the dataset and resource pages is confused and could be made clearer.

USABILITY ISSUES: FINDING A DATASET I

HOMEPAGE

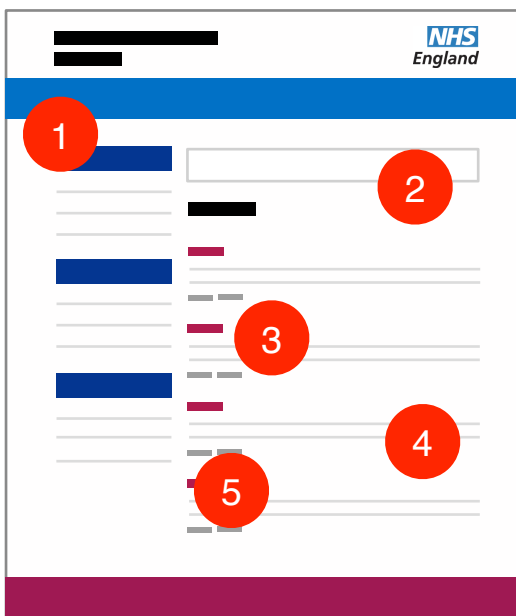


The homepage, where users started their search for a dataset. All users interacted with this page successfully, although some made negative comments on its visual design.

- 1 Many users complained about, and some shared screenshots of, not being able to see the icons for dashboard, settings and logout in the navigation bar.
- 2 Some people commented that group seemed a strange label for a collection of datasets.
- 3 As the startup representative commented, unless you're on the inside, it's hard to tell what sort of data you can find within it.
- 4 The term "CKAN statistics" is meaningless for most users.
- 5 The visual design is mundane and yet strangely fussy. For example, the search box has lots of visual noise in the form of many different coloured borders around it. The typeface used—Arial—has a less modern and sophisticated feel than Frutiger which is now used on the main NHS England site.

USABILITY ISSUES: FINDING A DATASET II

SEARCH RESULTS

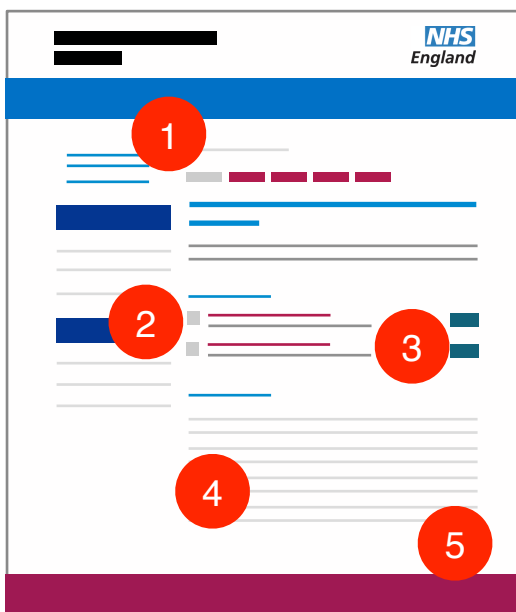


After searching on the homepage, users are taken to a search results page. Users were mostly able to interact with this page, with some confusion around contextual search and lack of date information and filtering.

- 1 Coverage date is one of the most important pieces of information by which people look for a dataset, but there is no filter available to narrow results using it.
- 2 When on a publisher or group page, it is not obvious enough that the search is now contextual. This confused the one or two users who found themselves in this situation.
- 3 Search results do not show coverage or release date unless it is included in the dataset title. This is a particular problem given that this is one of the key needs of analysts.
- 4 Different temporal releases are spread over many datasets, making the latest release harder to find, particularly in the absence of date filters.
- 5 Private datasets are not included in dataset search results. This confused all the users who attempted to find a private dataset as searching is the primary mechanism for finding datasets in the catalogue.

USABILITY ISSUES: FINDING A DATASET III

DATASET

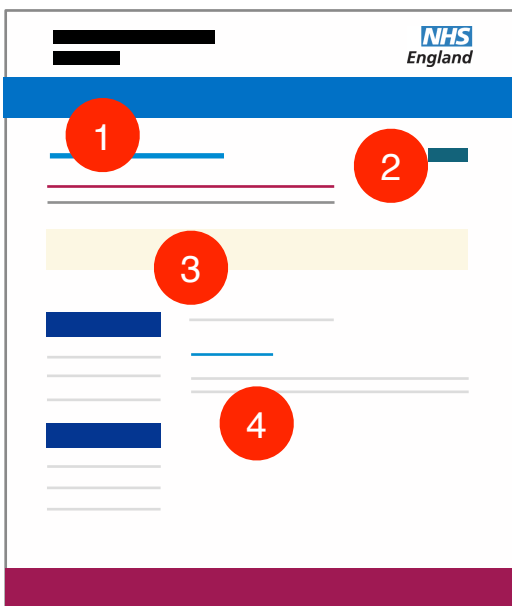


After clicking on a search result, users are taken to the page for the dataset. This page had issues with its layout and with the actions users take on individual files.

- 1 The visual hierarchy at the top of the page is confused: there is a general lack of clarity in the layout of the information.
- 2 The icons used to denote the file type of a dataset resource lack colour and as a result are much harder to distinguish from each other than they should be.
- 3 The “Explore” button on a dataset unnecessarily hides actions behind an extra layer of interaction. In addition, these actions change for different types of files in a way which is confusing for some users.
- 4 Metadata often includes redundant fields such as “Coverage Start Date” and “coverage_beginning_date”.
- 5 There is no way of getting in touch with the dataset owner and often no details of who they are. This was requested by one of the operations participant and the startup participant.

USABILITY ISSUES: FINDING A DATASET IV

RESOURCE



Finally, a user is taken to the page for an individual resource if they choose to preview it or click on its name. This page suffers from a lack of visual coherence.

- 1 The layout of the elements at the top of the page is messy and some—for example the embedded URL—appear to be superfluous.
- 2 The download button label varies between types of dataset from “Download” to “Get resource”. Some datasets show two differently presented download buttons on the same page.
- 3 Previews of files often do not load, and show overly technical error messages, for example “No handler defined for data type: pdf.”
- 4 Metadata often includes irrelevant fields such as “hash” and “resource group id” that give the impression that you are seeing too much of the system’s internal workings.

USABILITY ISSUES: ADDING A DATASET I

DATASETS



The datasets page, where most users started their attempt to add a dataset. This page is worked well for all users once they had found it.

- 1
- Most users commented about not being sure where to look to add a dataset, but all of them found it.

USABILITY ISSUES: ADDING A DATASET II

CREATE DATASET

The screenshot shows a form titled 'CREATE DATASET' with the NHS England logo. The form includes several input fields. Five red circles with numbers 1 through 5 highlight specific usability issues:

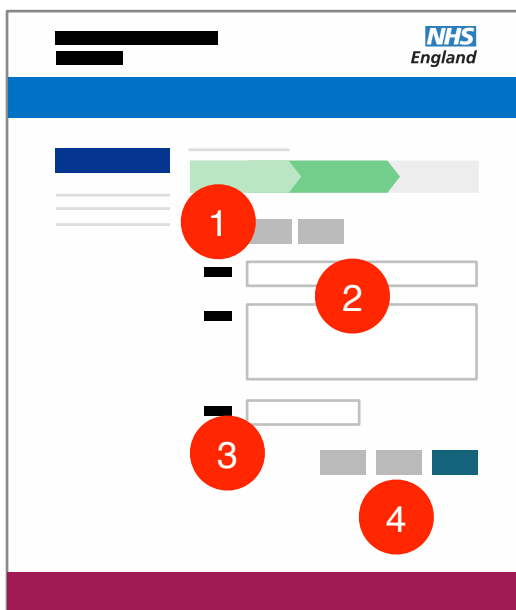
- 1: A redundant information box at the top left.
- 2: Unclear title and description fields.
- 3: Disappearing prompts for tags and dates.
- 4: Illogical order of options in frequency fields.
- 5: Unclear licence options.

This is the first page seen by a user once they start creating a dataset. The page does not do enough to help users through the process.

- 1 This information box at the top left appears to be entirely redundant: none of the users appeared to look at this. Information content like this is better provided in reaction to a user action when they are more likely to read it.
- 2 Four users commented that they were not sure what to enter into the title and description fields.
- 3 The helpful prompts in the fields for tags and dates disappear as soon as you start typing making it hard to know what format you should use at the point when you need this guidance most.
- 4 The actual order of options in the frequency fields does not reflect the logical order for these different options.
- 5 Many users commented that they were not sure what the different licences were or what they were had permission to choose. The startup participant commented that without a licence specified, a dataset was effectively useless.

USABILITY ISSUES: ADDING A DATASET III

ADD DATA



This is the second page seen by a user in the dataset creation process. Many users were left confused by what was happening on this page.

- 1 The non-standard file selection interaction was confusing to many users: once a file has been selected to upload, you cannot see the name of the file you have chosen.
- 2 Four users said that they were confused why they were being asked for a name and description for the file when they had already named and described the dataset.
- 3 Half of the analysts interviewed asked why it could not automatically detect the format of the file.
- 4 The “Save & add another” action at the bottom of the page is a qualitatively different kind of action from the other two presented and potentially confusing.

AREAS OF IMPROVEMENT

OVERVIEW

Base on this problem analysis, I have identified four potential areas for improvement where the rest of this work could focus.

SEARCHING AND DATES

Search result page design changes and other recommendations to make it easier for people to find datasets by coverage date.

ADD A DATASET WORKFLOW

Design changes and other recommendations for the add a dataset workflow—in particular the addition of in-context notes and guidance in the form—to give users more support when adding a dataset.

DATASETS AND RESOURCES

Design enhancements to the dataset page and the resource page to clear up the lack of clarity on these two pages and to eliminate the issues with the action buttons.

HOMEPAGE REDESIGN

A visual redesign of the homepage to make the site look more modern and focused while keeping within NHS England style guides.



FOR

NHS ENGLAND

AUTHOR

Mike Thompson

NHS REFERENCE

Paul Gavin