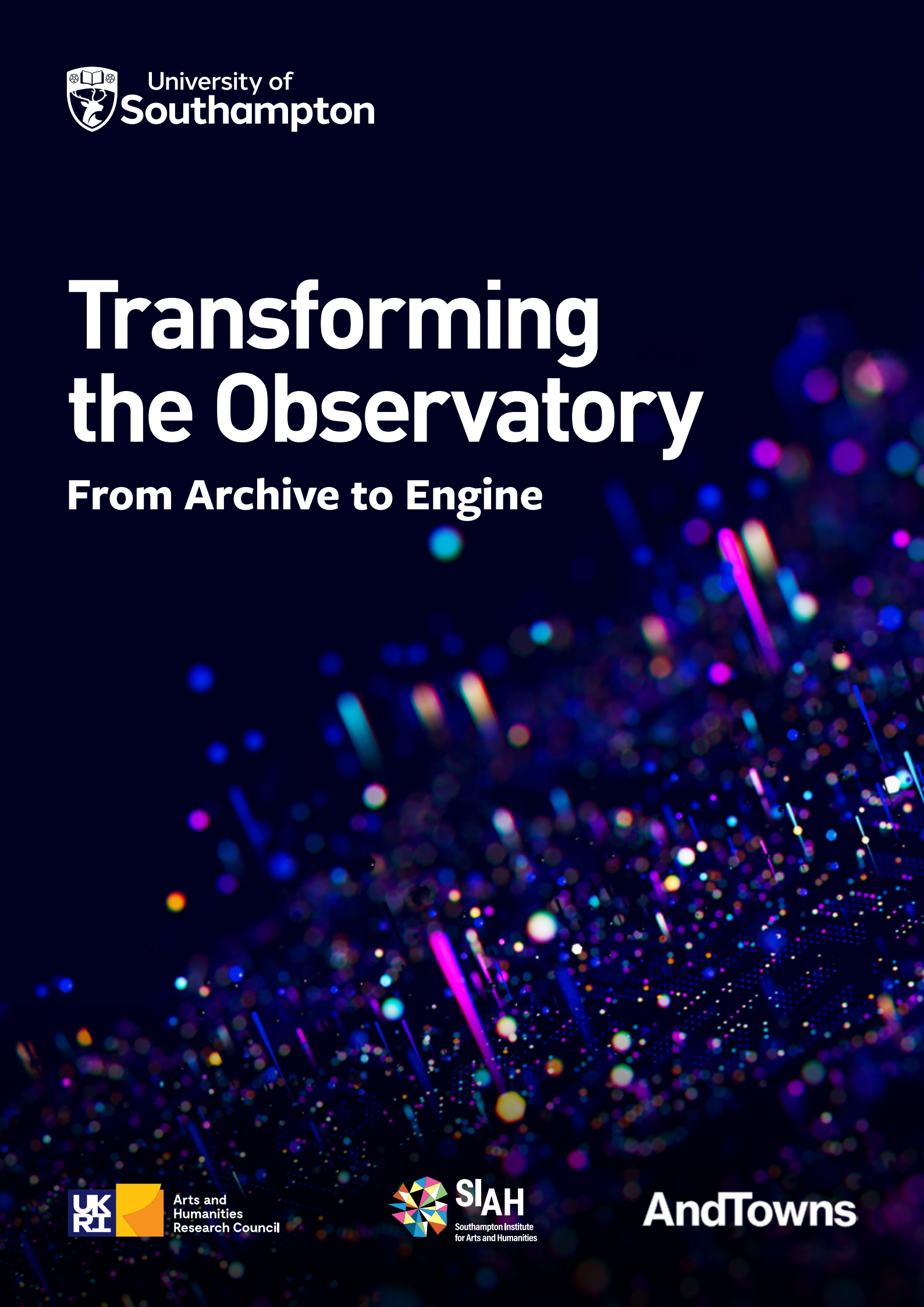
University of Southampton

Transforming the Observatory

From Archive to Engine



UKRI Arts and Humanities Research Council

SIAH Southampton Institute for Arts and Humanities

AndTowns

CONTENTS

INTRODUCTION

Transforming the Observatory is the third report from **Neighbouring Data.**

The first report, **Opening the Observatory**, outlines interim findings about the generation, connection and use of qualitative data in place-based decision-making. It focuses on the challenge of storing and analysing qualitative data, specifically through existing models of data observatories (Owen et al., 2023).

The second report, **Understanding the Observatory**, analyses the content and architecture of observatory portals. It purposively samples twenty data observatories with a review of the critical literature, focusing on three main subject areas: environment, technology, and place (Owen, Marsh, and Ashton, 2025).

This third report connects the findings from our previous reports with findings from a three-part research collaboration that included:

1. Part One: commissioning creative responses and hosting a workshop with 5 creative practitioners (June–October 2023).

2. Part Two: consulting with 25 professional data users on their use and aspirations for qualitative data (December 2023).

3. Part Three: developing prototypes for the Neighbourhood Insight Engine in line with our workshop and research findings (January–December 2024).

KEY FINDINGS

Qualitative data is widely used, and it is understood as a fundamental tool for providing engaged and meaningful place-based decision-making. Its insights are highly valued and cannot be replicated through other routes. It allows organisations to connect with and across communities in important ways by illuminating existing, tangible democratic deficits.

Yet there are ethical and practical concerns about the provenance, governance, accessibility, ownership, and use of qualitative data. There is a desire for a platform and protocols that will provide organisations and communities with the adequate tools to navigate this data. Such infrastructure needs to be reflexive and transparent when using this data but avoid making it homogenous, static, or unaffordable.

Part One: Working with Creative Practitioners

Method: We commissioned five creative practitioners to develop methods for representing qualitative data and to identify the challenges and opportunities involved in this work. We sought to explore ways to aggregate and render meaningful lived experience on place for community decision-making, and to understand how communities could expand on these forms of qualitative data.

Findings: Qualitative data takes many forms and can capture the richness of people's felt responses to the places where they live, work and volunteer. This data offers vital opportunities for local organisations and decision-makers to understand their communities in new ways. It allows users to draw connections across different types of experience in ways that cannot be replicated by other forms of data.

Challenges: Yet translating community engagement into qualitative data was deemed ethically and practically fraught, particularly when it was activated for use on a data observatory. Storing and accessing data gathered through community and creative practice must reflect the processes of participation and co-creation. It is important that the communities who produce data can also own and access that data, which mitigates the risk of data generation that is extractive and transactional.

Part Two: Consulting Data Users

Method: We held a workshop with twenty-five professional data users from a range of civic organisations. It sought to understand how these stakeholders were using qualitative data and what they considered to be the opportunities, risks and needs.

Findings: Organisations hold a significant amount of qualitative data, gathered through both formal and informal sources, ranging from surveys to social media feedback. They understand the value of this data and use it for a wide range of purposes. These functions include better understanding the communities that they serve, especially under-represented groups; adding complexity and nuance to quantitative data; effective advocacy, lobbying and storytelling; increasing appetite for democratic engagement; and demonstrating to communities that they have been listened to.

Challenges: There are concerns about how qualitative data can be gathered and used in inclusive, representative and rigorous ways that address assumptions and biases. There are also concerns about how data can be ethically, consistently stored and how access can be made easy, inclusive and cost-effective. There are many ad hoc processes in place, but users felt that better guidance and protocols could improve organisational practice. The informal ways of gathering qualitative data risked exacerbating rather than addressing existing democratic deficits.

Part Three: Neighbourhood Insight Engine

Method: Using the findings from the creative and civic workshops, we worked with practitioners and software engineers to better understand the issues and protocols of using qualitative data. It produced a deliberative tool that can be populated by different types of qualitative data and navigate a range of place-based projects. There have been so far two iterations of the Neighbourhood Insight Engine (NIE). **NIE 1.0** was designed by Joey Jones for the creative workshop, as outlined in Part One. NIE 2.0 was co-designed by Joey Jones, University of Southampton researchers in Electronic and Computer Sciences, and the Neighb ouring Data research team, as outlined

Findings: The Neighbourhood Insight Engine is a multi-user, hypertextual, offline, browser-based tool for exploring qualitative data that enables users to generate reports. This is a conceptual and practical move from an observatory, which holds or connects data, to a platform that enables users to supply and share qualitative data and which supports discussion, debate and decision-making. It creates a shared and level space (a "floor") to transparently connect diverse types of data. It provides the flexibility to work with diverse types of data. It creates the framework for conversations and decision-making using citizen-generated qualitative data.

Challenges: There is more work to be done to support this research, particularly in the provision of the skills, training and infrastructure that national bodies, local authorities and community organisations need to use these tools. Areas for development include workshop methodologies and guidance support for users.

PART ONE: EXPLORING QUALITATIVE DATA

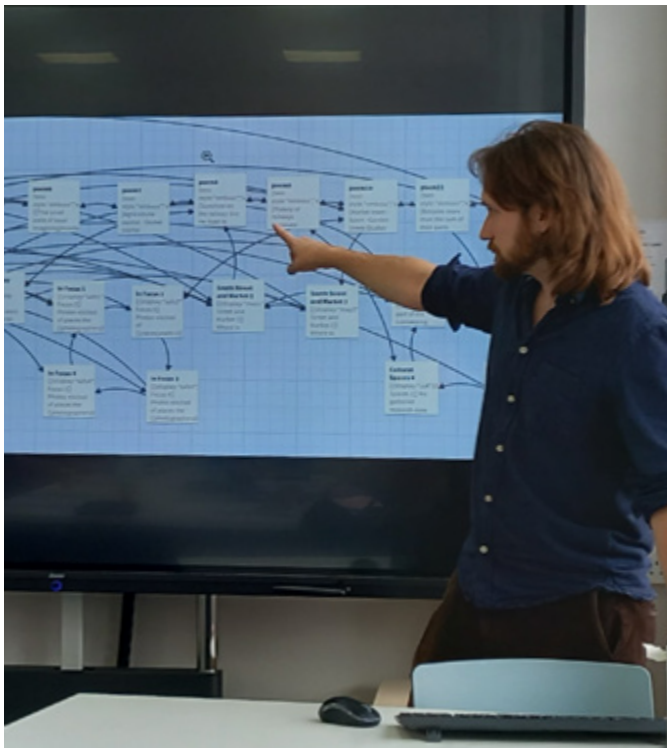
(June–October 2023)

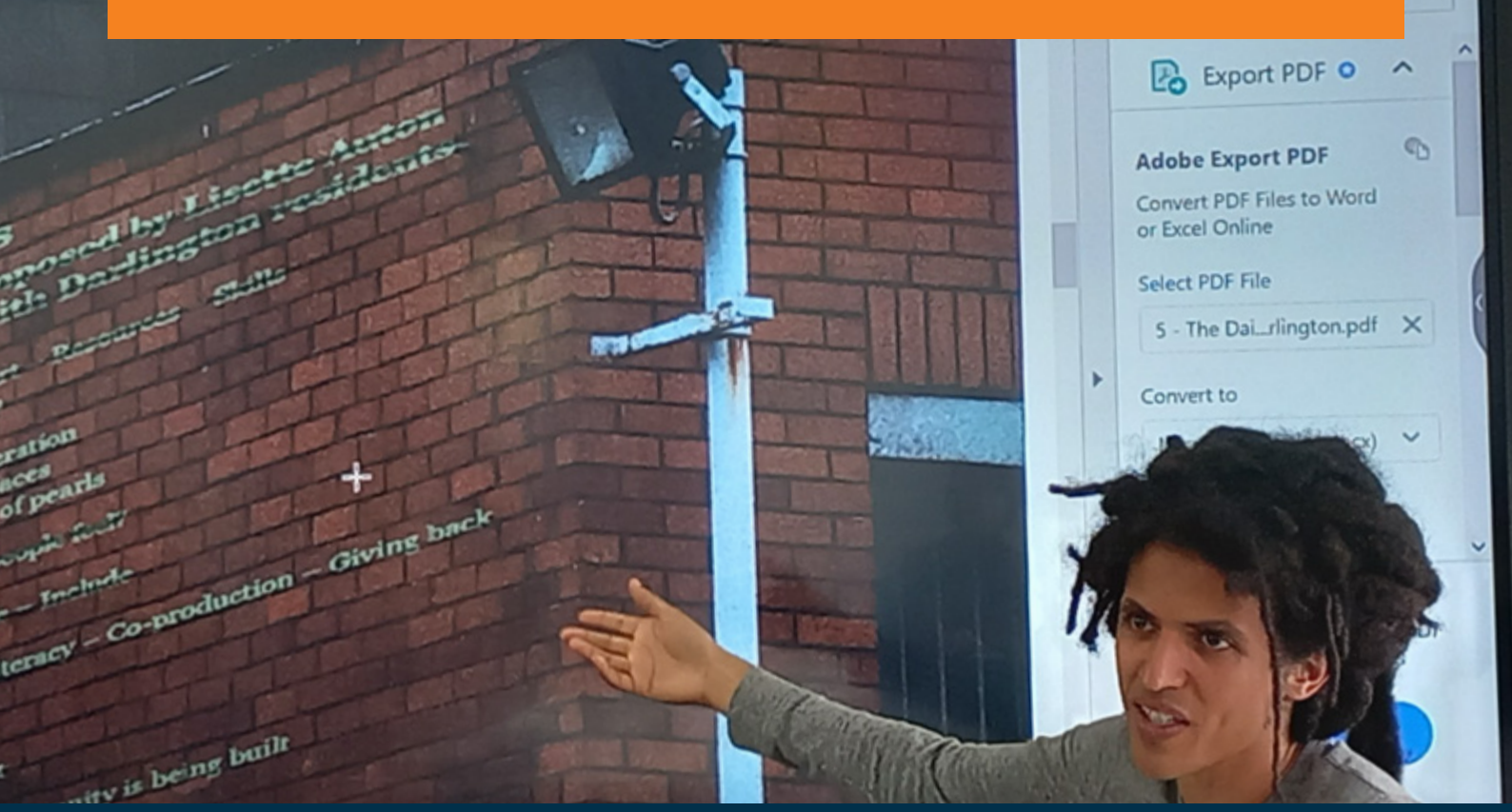
We commissioned five creative practitioners to explore a variety of methods for engaging with and representing qualitative data. The work sought to identify the practical concerns, ethical challenges and opportunities of working with this data. We focused specifically on the possibility of making connections between different kinds of creative, place-based data, as a form of qualitative data.

During the concluding workshop the practitioners presented demonstrations on:

1. an interactive 3D Minecraft map visualising data about a town that enables users to intervene in their streetscape by adding or demolishing buildings
2. an audio-visual piece illustrating different visions of a town's future that used coloured text and field recordings based on community responses
3. an explorable website, the Neighbourhood Insight Engine, with a game-like interface designed to prompt user curiosity about place-based data
4. five satirical online newspaper articles reporting on a fictional government initiative for twinning together towns in the UK
5. a roadmap for data observatories considering their infrastructure, stakeholders, ethics, sustainability and social impact

The presentations were followed by a facilitated conversation, addressing divergences between the participants' creative or scientific practice and the research questions. We organise the following discussion into five key analytical criteria that have been deployed across the **Neighbouring Data** reports: Audience, Visualisation, Narration, Use, Maintenance.





Audience: Who is qualitative data for? Who should access it?

Both the definitions of data and the nature of community engagement were widely contested. Qualitative data was not always considered the right term for how artists represented data from their work with communities. Visualisation and narration were regarded as forms that resisted aggregation outside of the context in which they emerged. Creative practitioners were not enamoured by the act of decontextualising data that they had elicited with communities and using it for comparison elsewhere.

Ethical concerns were raised about the risk of using creative methods in extractive and transactional ways. Data considered to be valuable should be obtained through primary work with communities and co-created over a long period of time. This community ownership of data was key to the productive collaboration between artists and communities. Why, how, and for whom data collected must be foregrounded in any approach. The need to be transparent with communities was another recurring sentiment. Meaningful consent is a crucial mechanism for producing credible qualitative data.

Visualisation: How can qualitative data be represented effectively? How should it be visualised?

Participants sought a visual representation of data that could encourage users to not only uncover information but illuminate complex community needs and significant policy challenges. As one creative practitioner put it, by using creative visualisations of qualitative data, 'you can discover more of what you didn't know you were looking for' (Creative Workshop, 2023). In this case, 'you're not just checking for isolated, measurable points but attempting to look at the whole picture' (Creative Workshop, 2023).

The five presentations demonstrated how different projects produce different kinds of media, artefacts and outputs, which creates different opportunities for connections and reflection. Bringing into dialogue diverse forms poses obvious challenges, but the interaction of testimony, maps, timelines, audio, images, and creative writing can engineer novel insights about places and the communities who inhabit places. While the forms of data may not be standardised, the conversation between different representations of people's experiences can support decision-making about shared assets and services.

Narration: What should accompany this data? What story should qualitative data be telling?

The workshop discussion supplied several interventions about how distinctive place histories influence storytelling about places. The narratives that people present about their places hold a certain truth value, regardless of their fidelity to actual events. Creative practitioners felt it important to recognise that the artist's skill lies in making community voices engaging and coherent, but that the artist must also resist editorialising and influencing this act of translation.

Yet there are paradoxes inherent to that perspective. The narration of people's lived and felt experience will inevitably feature the biases and assumptions of the narrator. It is important to reflexively recognise that the position and role of the creative practitioner is never neutral in the presentation of these community experiences. This byproduct of engagement should not prevent activities from taking place, rather it should inform the understanding and analysis of the evidence gathered from this work.

Use: How should users interact with qualitative data? What should be their experience of using this data?

One workshop participant argued that to understand and represent qualitative data without privileging the process would render the data 'meaningless in some ways' as such work is 'massively contextual' (Creative Workshop, 2023). Creative practitioners explicitly advocated for a methodological model through which they worked 'over long periods of time with a group to try and get to know it really well' (Creative Workshop, 2023). By increasing participation, this model provides a sustained form of engagement with seldom heard communities and young people, but it is less focused on the practical application of the qualitative data that is being generated.

As one practitioner put it, 'the art is the relationship that I hold with people in the moment' (Creative Workshop, 2023). This comment characterises community engagement as a dialogic practice, which poses a problem for reading qualitative data as an artwork. This point leads to questions about audience: there are intended and inadvertent consequences of this work and who it addresses. It is vital to draw out the tension between local and national concerns, especially when hyper-local issues are often framed in a national political context. One practitioner suggested that the UK media already provides a version of the qualitative data observatory, insofar as it narrates and elides community voices (Creative Workshop, 2023).

Maintenance: How can the representation of this data be updated, maintained and sustained?

The computer scientists who participated in the workshop were more interested in the challenges of data silos and open data than in community engagement. Web observatories share data and supply information, but citizens need to have a voice in how this dissemination occurs. Co-operative and curated datasets can contribute to a decentralised internet, but app developers must ask for explicit consent from users, who can withdraw and remain in control of their data. These models must enable citizens, businesses and artists to always have access to relevant data.

A **Personal Online Data Store**, for instance, is a mechanism that individual users can use and control. This type of data store can democratise how citizens access and engage with data. Such a mechanism could be funded as part of both national and local infrastructure. Data, in this case, is seen as a citizen asset. A data store provides citizens with a space to hold and share their digital resources, which can enable a more balanced dialogue with governments.

Neighbourhood Insight Engine

From the presentations and discussions, we selected a proposal that could best augment our understanding of how to connect, represent and analyse qualitative data. One presentation from the workshop, the **Neighbourhood Insight Engine,** took examples in the resource pack and connected them through a website portal. The engine invites users to take time with each example, navigate connections between them, and record insights and reflections as they traverse the pages.

The practitioner, a digital w, described it as a hypertextual approach (Creative Workshop, 2023). Creating a hypertext encourages the user to engage with all parts of a qualitative data collection to find a rich set of connections. The practitioner likened the approach to printing out sheets of paper, pasting them to a wall, and linking them together with string. This method takes advantage of the hypertext platform by enabling non-linear and non-hierarchical reading experiences. We took this approach to a civic workshop with key policymakers from across local government, national cultural organisations, and policy network groups. This event produced major reflections on the development Neighbourhood Insight Engine and broader insights on the creative workshop.

PART TWO: CONSULTING DATA USERS

(December 2023)

In our civic workshop, we consulted data users from a wide range of civic sectors about their current use of qualitative data and the roles and functions they think this data might serve in the future. The event included:

- **Session 1:** a presentation on the project's interim findings, including a session exploring the Neighbourhood Insight Engine.
- **Sessions 2 and 3:** interactive workshops designed to understand the issues, capacities and potential of using qualitative data to inform place-based decision-making. Each workshop was organised through a set of semi-structured questions, which organisations addressed in pairs and then as a group. The questions and their outcomes are included below.

Participants were drawn from:

Local Authorities:

Isle of Wight, Southampton, Bournemouth, Portsmouth, Eastleigh, Southend

National Organisations:

Historic England; Local Government Association; Department for Levelling Up, Housing, and Communities; Local Business Improvement Districts

Community Organisations:

No Limits, SO18 Big Local, Testlands, Southampton Voluntary Services

Session 1: The Neighbourhood Insight Engine

The Neighbourhood Insight Engine emerged from the creative workshop with practitioners and computer scientists. Its first iteration, NIE 1.0, was presented at the civic workshop by the lead designer Joey Jones. This version is accessible on the **University of Southampton website.**

The presentation drew on reflections from Tee Guidotti's article **The Observatory**, which considered the possibilities of a data observatory as a model for studies in health, society, and the environment. In this context, the observatory can:

> Allow the investigators to compile comparable events into a study set that can be used to compare commonalities and differences in experience, leading to more formal testable hypotheses and policy studies (2022: 829).

This entry point established that different projects and their associated data could be viewed as comparable events. The aim was to consider how to practically facilitate this comparison.

The NIE hypertext approach enables **interconnections between data for decision-making**. This framing follows Noah Wardrip-Fruin's line of questioning in **What hypertext is** (2004), which emphasises the links and connections between texts and how these interactions construct knowledge via associations. The NIE does not envision hypertext just as a simple means of linking data within an observatory structure. It understands and facilitates knowledge and insights through these associations. The NIE user can take time with data, navigate connections between them, and record their insights and reflections as they go.

The NIE resonated with participants who provided the following insights:

- The NIE focus on building associations and exploring insights over storing data was promising because data is often used on a case-by-case basis and the over-organisation of data can impose associations and meanings.
- The NIE can be used to show how decisions were reached. Rather than simply presenting findings or analysis, the relationships between data can elicit insights and make decision-making more transparent.
- The pressures of limited resources underpin an eagerness to use already existing data, instead of gathering new data, to understand different issues. Using data relating to one policy area has the potential to affect policy in another area.
- The open-ended, exploratory nature of the NIE embraces the challenge of whether different data generation (e.g. surveys) ask the right questions and can be activated for new purposes. The fit between question and purpose is less a limitation when the aim is to gather multiple perspectives, data sources and insights to address a policy problem.
- Participants shared an interest and investment in training and supporting citizen researchers. The NIE enables equitable connection and exploration of diverse datasets. This function means that robust, credible citizen data relates to other datasets on an equal basis as part of the decision-making process.

Participants raised several points that helped to iterate the NIE, as it is outlined in Part Three. These queries included:

- How the insights generated and recorded by users on the NIE might be engaged with and used by policymakers.
- How the NIE addresses long-standing challenges of qualitative data so that it enables inclusive representation, especially from seldom heard communities.
- How to ensure that NIE access is equal, that cost considerations are appropriate for the organisations, and that decision-making improvements can be identified.
- How to maintain open, co-production with communities in the development of the NIE.
- How to evaluate the costs of developing, understanding and engaging with the NIE, particularly for under-resourced organisations.

Session 2: Understanding Qualitative Data in Place

This session focused on understanding the current practice within organisations for gathering, analysing and using qualitative data. These are the questions and collated answers produced by the activity:

What forms and sources of qualitative data are currently used by your organisation?

- Formal interviews and focus groups with communities and service users.
- Informal sources included public engagement events, emails, canvassing, open box surveys, and social media (including human-AI interactions).
- Co-produced creative outputs, including from formal consultations.
- Observation and audience research through surveys, social media, local and national media, and cultural artefacts.

How does your organisation currently organise this qualitative data?

- Through formal monitoring collated by year and/or by organisation and collected as digital forms on laptops.
- Through quantitative structures (e.g. Dates and Times, Version Histories) that are given thematic qualitative groupings and platform names.
- Through qualitative software (e.g. NVivo, Excel, Word, Digital Folders).
- Yet the value of these modes is qualified by some users and not every organisation is equipped to use these methods consistently, if at all.

How and when does your organisation currently use qualitative data?

- Through inclusion in published reports (e.g. blogs, academic papers, creative projects, and artefacts) but further use relies on corporate memory.
- On a case-by-case basis to understand the significance and effectiveness of specific interventions, but this data is not stored and there is little way of tracking findings longitudinally.
- To illustrate, narrate and discuss quantitative data (e.g. as pictures, on social media, through case studies).
- To connect with communities by anonymously sharing real-world examples of available support given to other residents, and to give feedback to volunteers who can shape decisions and provide insights into community responses.
- For political advocacy: qualitative data is an emotive hook to persuade financial decision-makers who otherwise require quantitative data for business cases, funding applications, and forward planning.
- To understand issues that cross silos within and between organisations (e.g., culture, wellbeing, and public health).
- Through a large-scale corpus analysis: qualitative methods and insights can unveil themes and shape deep-dives into subject areas.

Conclusion

- **Provenance and rigour:** Qualitative data can seem cheap and ubiquitous, but it is difficult to make sure that this data is being approached with focus, rigour, and credibility. There are serious concerns about the position and bias of organisations in both the gathering and analysis of this data.
- **Governance and practice:** Although organisations face what they describe as a surfeit of qualitative data, there are serious questions about the ethics and consistency of eliciting, storing and accessing this data. Participants felt that the many ad hoc methods in place for organisations could be improved by better guidance and protocols.
- **Inclusion and representativeness:** Qualitative data collection is often targeted as it is based on quantitative data or previous qualitative research (e.g. experience of hate crime in a particular neighbourhood). Yet it is often difficult to ensure that this data is representative and that the views of seldom heard communities are included. The informal ways of gathering qualitative data risked exacerbating rather than addressing existing democratic deficits.

Session 3: Aspirations for Qualitative Decision-Making

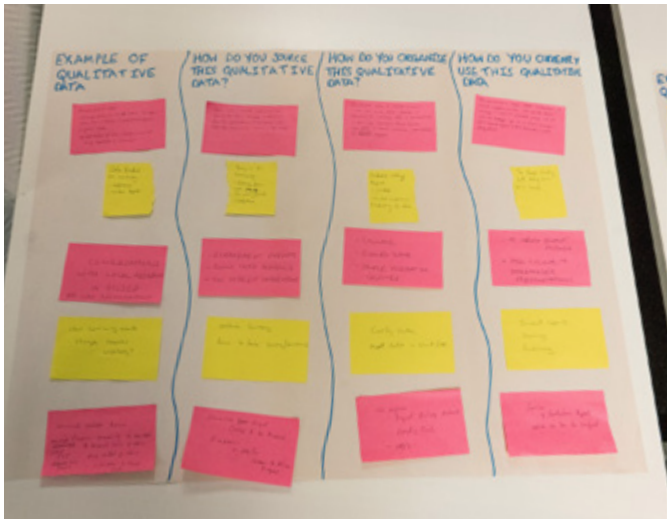
This session focused on understanding participants' future aspirations for qualitative data. It was specifically interested in what organisations hoped qualitative data might achieve and how they would ideally like to work with this data. These are the questions and collated answers produced by the activity:

How would your organisation like to source qualitative data in the future?

- Consistently, regularly and from a cross-section of the community, including on-site and informal conversations that can produce genuine community engagement and real, defined outcomes.
- As a form of assessment from uninvolved individuals who have no inherent investment in what is being discussed.
- Through the training of citizen peer researchers to write reports rather than only gather data.
- Quickly and openly from online and accessible platforms.

How would your organisation like to organise qualitative data?

- Data to be localised in a calibrated system, with adequate data regulation.
- Regular inputs of data organised in an accessible manner that can be used easily for a variety of purposes.
- Data to be made as easy as possible to navigate, especially for volunteers.
- Data to be aggregated in an order that can provide visual and summary data without losing the qualitative element.
- Data to be organised in ways that are coherent and unbiased, so that associations can be made across different datasets and questions, and which do not rely on human decision-making in how the data is grouped.
- Data to be in an accessible format that can be readily shared if needed.

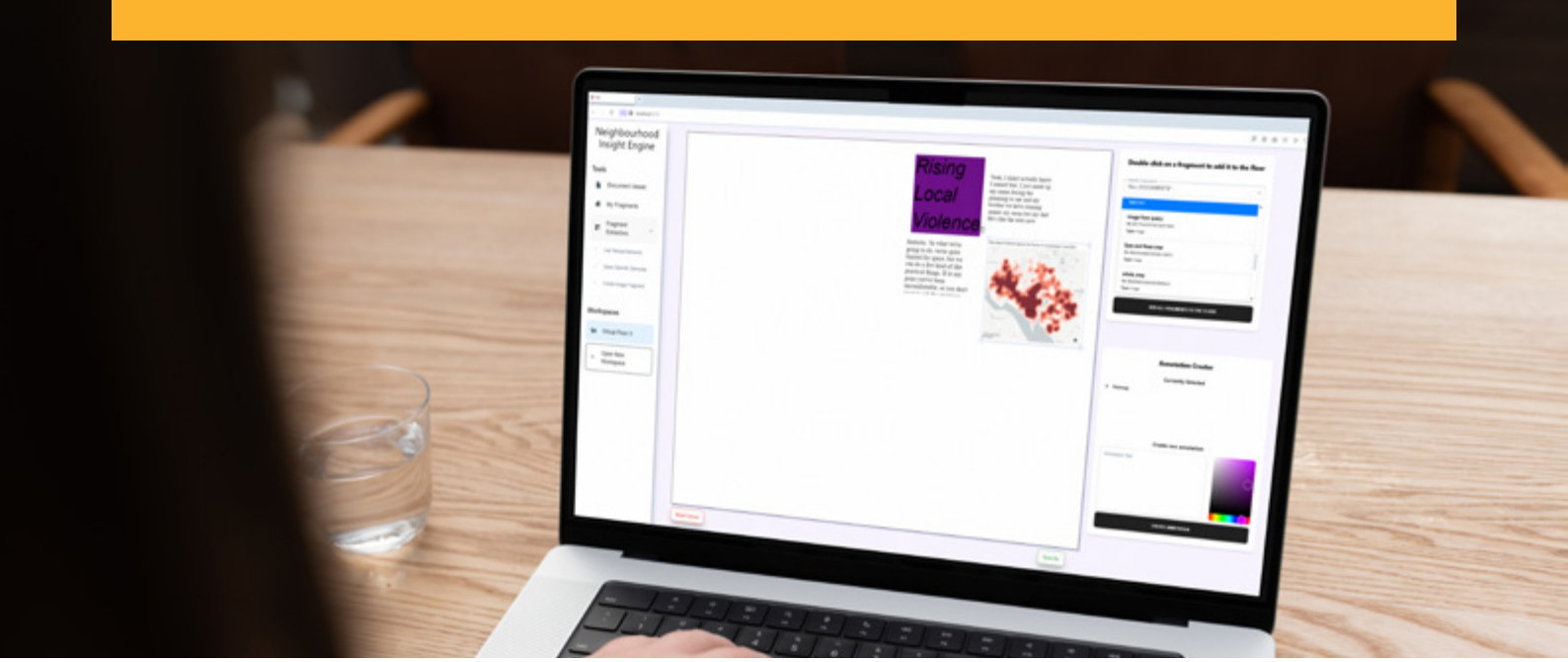


How would you like to use your qualitative data?

- To inform key decisions for community-focused delivery and provide evidence for planning community interventions (e.g. museum exhibition learning, events, and community outreach programmes).
- To allow all people involved in the decision to have a voice expressing their feelings and preferences.
- To address a democratic deficit in decision-making and demonstrate to communities that they have been meaningfully heard.
- To make decisions that inform creative choices but are generally backed by human interaction, and to ask better questions to define what is not known.
- To ensure a consistency over time that can meaningfully counter quantitative demands (e.g. monitoring reports that ask: "How many people are volunteers or in work?" but give little space to capture a qualitative understanding of what is meaningful volunteering or work).
- As an assessment tool to quantify the value of a project, and the success or failure of the activity.
- To facilitate a second, granular sift of grant applications which follows a first, quantitative sift.
- To shape a consultation process and to inform a final report.
- To illustrate quantitative data (e.g. as a layer on a map).

Conclusion

- **Qualitative data is uniquely important:** Existing reporting and decision-making tools place too much emphasis on quantitative data. The value of qualitative data for understanding communities is felt and understood across sectors.
- **Qualitative data has multiple and nuanced roles:** It supports advocacy, communicates across a democratic deficit, identifies emerging and less understood issues, and complements quantitative data by providing a richer evaluative oversight.
- **Qualitative data is understood to be complex and dissonant:** It reveals contradictions that organisations understand and respect (e.g. between ensuring safety and enabling access with respect to a local service or asset).
- **There is insufficient training:** People and communities need skills to collect data, so that the outcomes are not the product of tick-box consultation and personal preference but that of robust research.



PART THREE: THE NEIGHBOURHOOD INSIGHT ENGINE

The findings from Part One and Two were used to explore the potential of the Neighbourhood Insight Engine for connecting qualitative data and decision-making. The NIE was designed to help organisations use qualitative data in a rigorous, credible, and reflective way. It continues to develop as a platform for exploring and connecting data.

NIE 2.0 is a multi-user, hypertextual, offline, browser-based tool for exploring qualitative data that enables users to generate reports. This is a conceptual and practical move from an observatory, which holds or connects data, to a platform that enables users to supply and share qualitative data and which supports discussion, debate and decision-making.

Building on the creative and civic workshops, NIE 2.0 was developed in a three-month phase, bringing together the Neighbouring Data team (Marsh, Ashton, Owen) with lead-designer (Jones) and colleagues from Electronic and Computer Sciences (Dr Richard Gomer and Owen Richards) at the University of Southampton.

Appendix A includes the full technical report for NIE 2.0. The following section summarises the key features:

- The development stage envisioned NIE 2.0 as Computer Assisted Qualitative Data Analysis Software for exploring and thematically coding qualitative data.
- Extracting fragments of a document is key to thematic analysis, so the NIE provides multiple ways to do this.
- The primary use case for the fragments during the development stage were local government surveys and image-based responses (e.g. photos, maps, timelines) included in the **Neighbouring Data Resource Pack.**
- The **My Fragments** window gives the user a place to view all fragments that they have extracted from the documents uploaded. They can be filtered by the document from which they were extracted. Clicking on a fragment opens a panel that shows a preview of the fragment, a list of all annotations linked to this fragment, and has space for extra functionality in future, such as the ability to edit data about the fragment.
- The **Virtual Floor** view enables users to arrange these fragments and add annotations to emulate the thematic analysis process known as coding. These fragments and annotations can be effectively arranged into piles, helping to demonstrate ideas the user has inferred from the qualitative data.
- When selecting fragments, both individually and in groups, a window appears at the bottom-right for creating text annotations. The annotations can be given a colour to help distinguish concepts, which is, again, similar to coding in traditional qualitative analysis.

NIE 2.0 set out in Appendix A is informed by the Neighbouring Data reports in the following ways:

→ ***Opening the Observatory*** addressed the importance of connecting data at different scales, including the hyper-local neighbourhood scale. NIE 2.0 enables this scale of data to join with other data on the same virtual floor without privileging certain types of data or introducing data hierarchies.

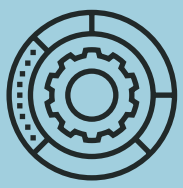
NIE 2.0 creates a shared and level floor to transparently connect diverse types of data.

→ The creative workshop (October 2023) amplified the importance of ongoing community engagement with qualitative data that reflected its generation and creative forms. NIE 2.0 respects the many forms that qualitative data can take while allowing connections between them. It enables this function as users can upload and view documents in .html formats from a variety of data sources: survey responses, interview transcripts, images, and more.

NIE 2.0 provides the flexibility to work with diverse types of data.

→ **The civic workshop** (December 2023) foregrounded the importance of consistent, shareable qualitative data-driven storytelling. Participants emphasised the need to include community voices and to use citizen data in decision-making. While demonstrating transparent and reflexive uses of data, NIE 2.0 contributes to the exploration of citizen and group decision-making. This approach could encompass citizen assembles and citizen juries (e.g. Local Government Association and Citizens in Power). As David Jubb writes in an essay on citizen-led decision-making, 'it's not just who makes decisions, it's how they're made' (2024).

NIE 2.0 creates the framework for discussions and decision-making through citizen-generated qualitative data.

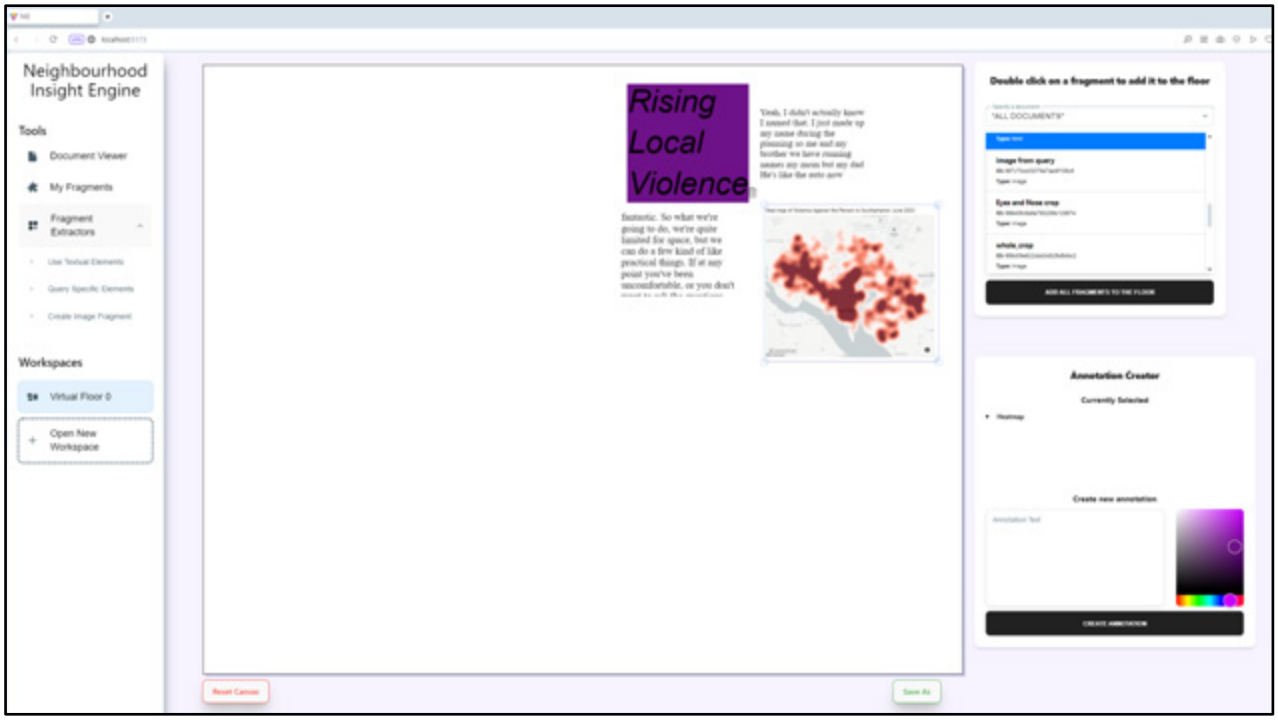


From Archive to Engine

Qualitative datasets are often large and heterogenous. The Neighbourhood Insight Engine offers a model for connecting, discussing and interpreting various types of creative, place-based qualitative data. It allows qualitative data to be activated in different contexts and circumstances through a two-directional platform of engagement and participation. This platform can draw together community perspectives and specific data units to generate local place narratives. It is these functions that transform the observatory from an archive to an engine, and which constitute the focus and argument of this report.

Neighbourhood Insight Engine

Technical Progress Report



A report detailing the features, technical approach and next steps, for the Neighbourhood Insight Engine's prototype.

Produced by Owen Richards - 31/08/2024

Table of Contents

Overview

The Neighbourhood Insight Engine (NIE) is a tool aimed for gaining insight from collections of qualitative data. It was envisioned as software for exploring and thematically coding qualitative data, with the primary use case being local government surveys, and image-based responses (photos, maps, timelines etc.). The NIE was designed as Computer Assisted Qualitative Data Analysis Software (QDAS or **CAQDAS**), and as such has features that elevate it above existing offerings (such as NVivo, and Atlas.ti).

Extracting fragments of a document is key to thematic analysis, so the NIE provides multiple ways to do this. The system then allows you to arrange these fragments in a view known as the virtual floor, adding annotations to emulate a process known as 'coding'. These fragments and annotations can then be effectively arranged into piles, helping to demonstrate ideas the user has inferred from the qualitative data.

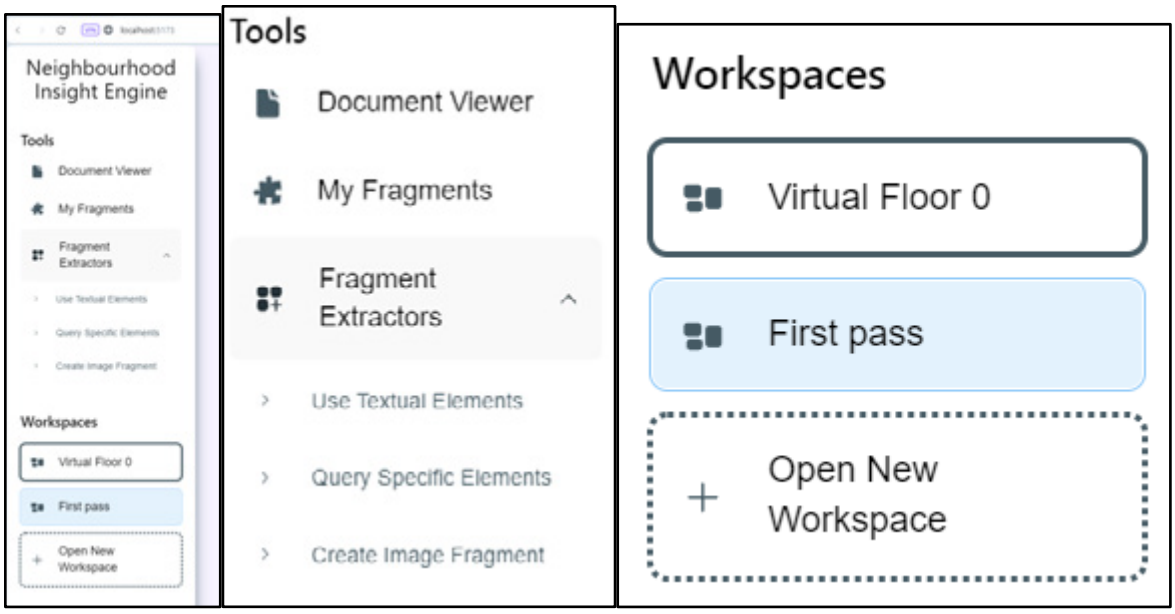
The NIE has been envisioned as a web-app. The app is built using Vite + React for JavaScript, two popular frameworks for web development that are likely to continue to have long term support.

This initial prototype is the result of three months development by a computer science graduate for the AHRC-funded Neighbouring Data research, part of the And Towns projects. This work was undertaken in collaboration with the Web Science Institute and Southampton Institute for Arts and Humanities at the University of Southampton.

Features/Functionality

Sidebar

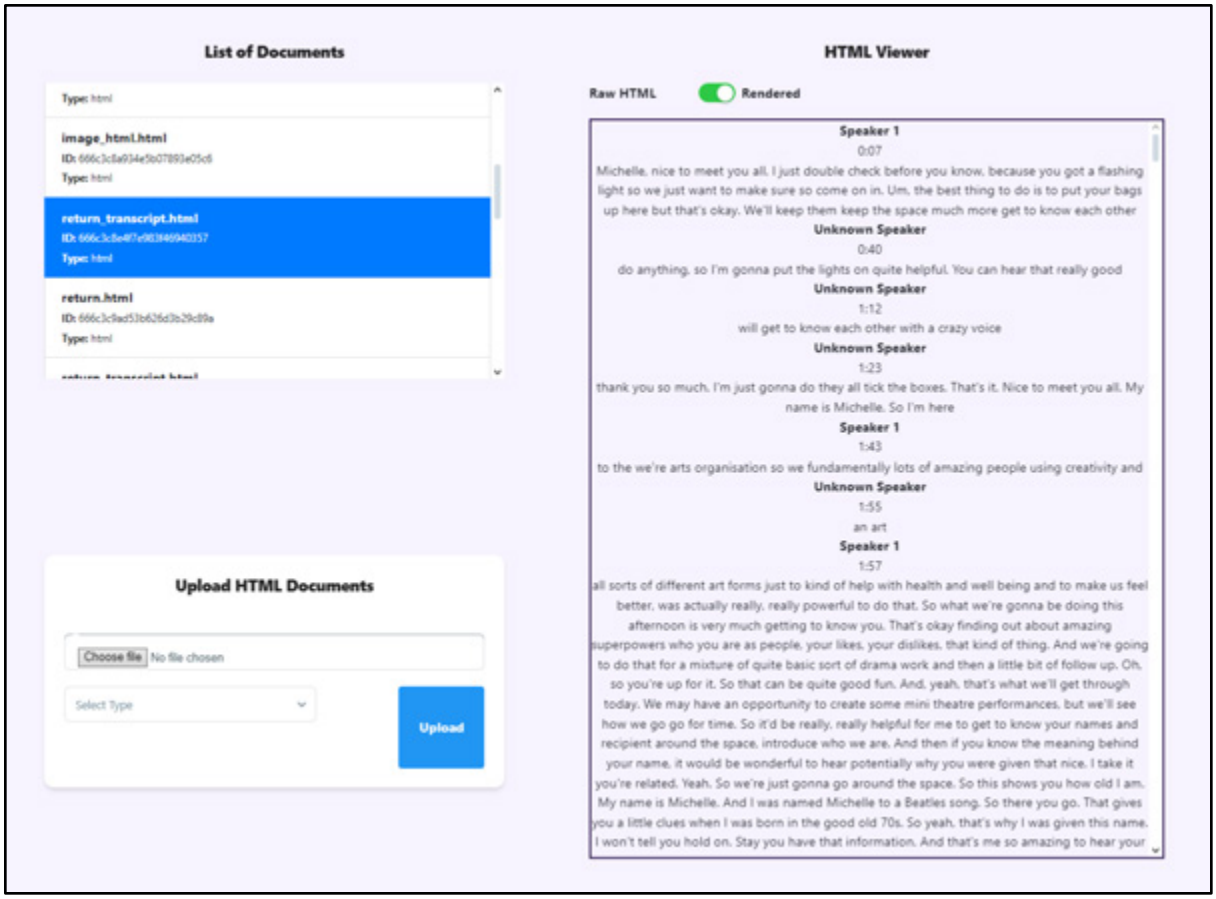
To navigate between the different sections of the NIE, a **Sidebar** persists on the left hand side of the screen. It offers links to various features of the app which can be clicked on to launch them in the main area of the screen. The selection is split into tools and workspaces. Tools are opened one at a time, whereas workspaces act as multiple tabs, for opening multiple virtual floors or documents simultaneously.



*Screenshots of the **Sidebar**, along with close-ups of the specific options available to users.*

Document Viewer

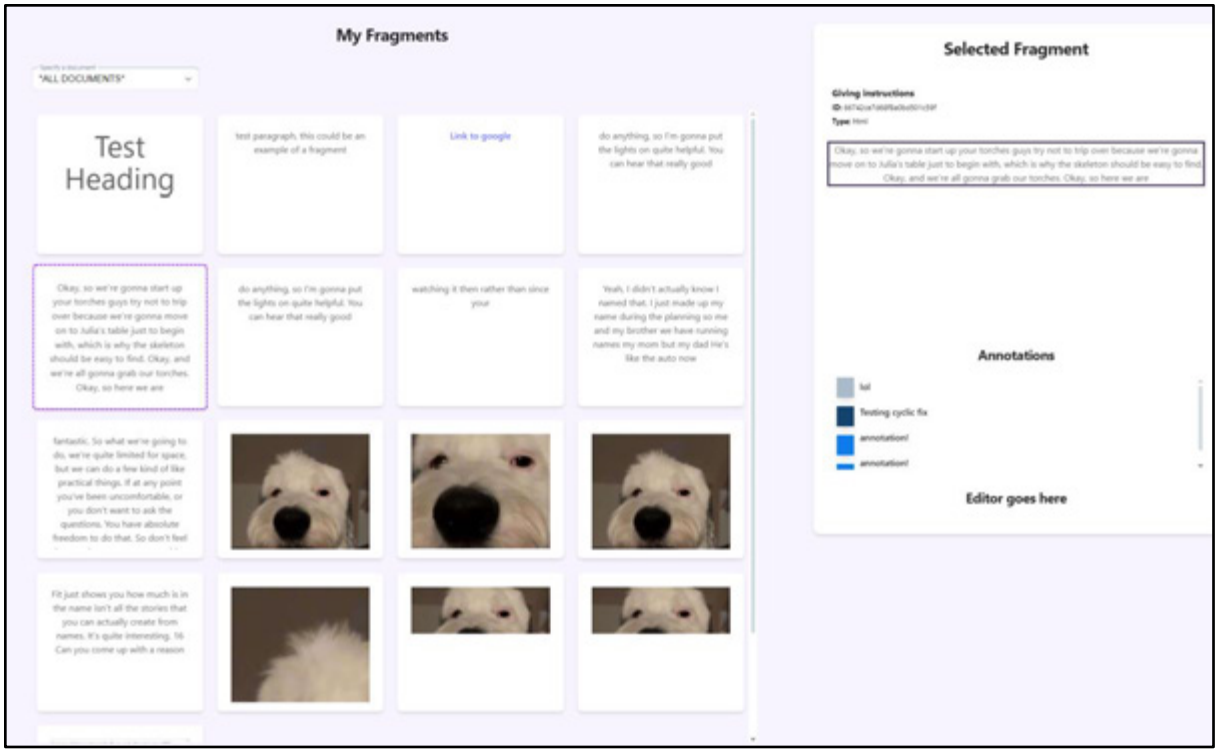
The system allows users to upload and view documents in .html formats. These can take the form of survey responses, interview transcript, images and more. As well as viewing crucial data about the document, the user can choose to view the document's raw HTML content, or a rendered view, similar to how it will look if embedded in a webpage.



Screenshot of the Document Viewer in use

My Fragments

This window gives the user a place to view all fragments that they have extracted from the various documents uploaded so far. They can be filtered by the document which they were extracted from, using the drop-down at the top of the list. Clicking on a fragment opens a panel on the right-hand side, showing a preview of the fragment, a list of all annotations linked to this fragment, and has space for extra functionality in future, such as the ability to edit data about the fragment; delete the fragment etc.



Screenshot of the My Fragments view in use.

Fragment Extractors

There are multiple tools available for extracting fragments. A drop-down in the **Sidebar** allows you to select each tool and open it in the main area. In each of these cases, the new fragment can be named by the user; future functionality to add tags to the fragment is also supported here but currently serves no purpose. For images, there is the **Create Image Fragment** tool in which a specific region of the image can be cropped for use as a fragment.

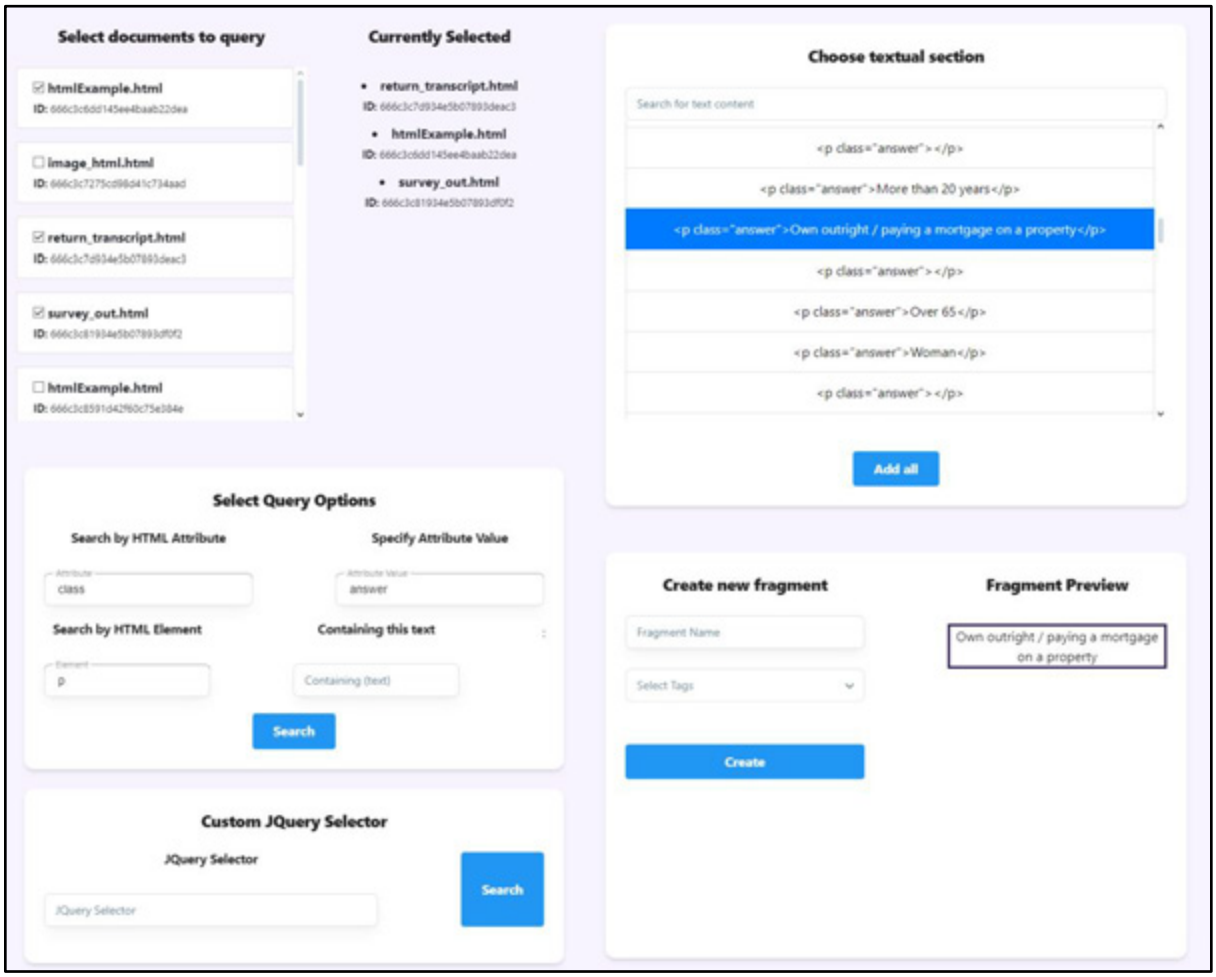
Screenshot showing the Create Image Fragment tool's cropping functionality.

For textual fragments, or other html sections, two views are provided. The **Query Specific Elements** tool provides two contrasting approaches to html querying. JQuery selector statements can be entered to query the HTML documents, which are then parsed by the system, producing the results in the window at the top-right. JQuery is a long-standing and well documented tool for manipulating HTML documents, so advanced users more familiar with qualitative analysis will appreciate this feature.

For those unfamiliar with JQuery, but who still want to query using HTML features, there is support for specifying attributes, elements and their values. The system creates custom JQuery selection statements under the hood to support this. The example below shows a user finding the answers to each question in a survey response using a simple query.

The document selector at the top-left allows you to select multiple documents, all of which will be queried, allowing the results to be compiled together. This could be useful when searching through multiple survey responses etc.

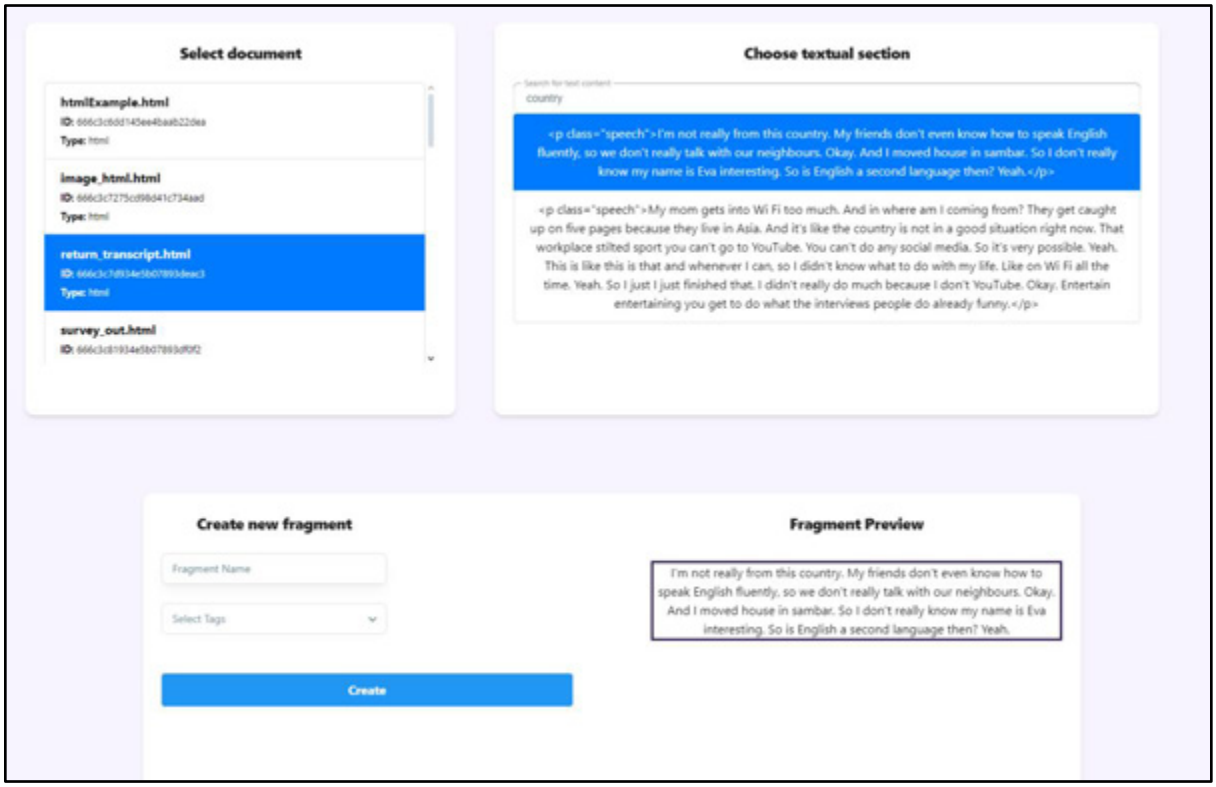
Double clicking on a result in the scrollable results window will preview the fragment in a similar way to the **Create Image Fragment** tool. These results can also be searched for specific terms in the HTML content.



*Screenshot showing the **Query Specific Elements** tool finding and selecting search results.*

The **Use Textual Elements** tool is used for more simplified searches. The user can search for textual content within the html and select these sections specifically. It will list all possible

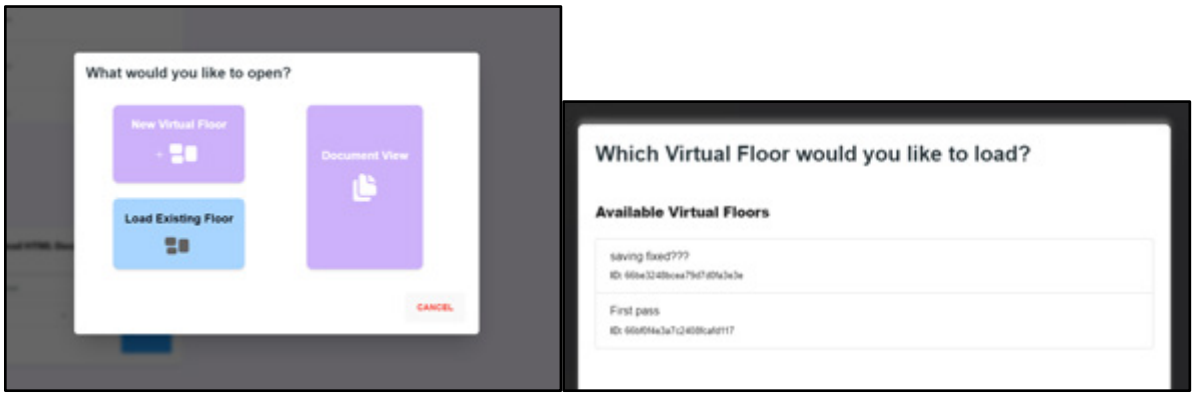
HTML fragments containing text before a search term is entered. This is more tailored towards novice users, and those less familiar with HTML's linguistic features, such as elements and attributes, however it should also prove useful for advanced users who need to quickly find related ideas within a document.



Screenshot showing the **Use Textual *Elements*** *tool finding and selecting search results.*

Virtual Floors & Workspaces

A button labelled **Open New Workspace** at the bottom of the **Sidebar** brings up a pop-up menu, allowing the user to select either a new/existing virtual floor, or to open a document *(not yet implemented)* for easy viewing. Selecting an option here will open a new tab for the selected workspace in the **Sidebar**.



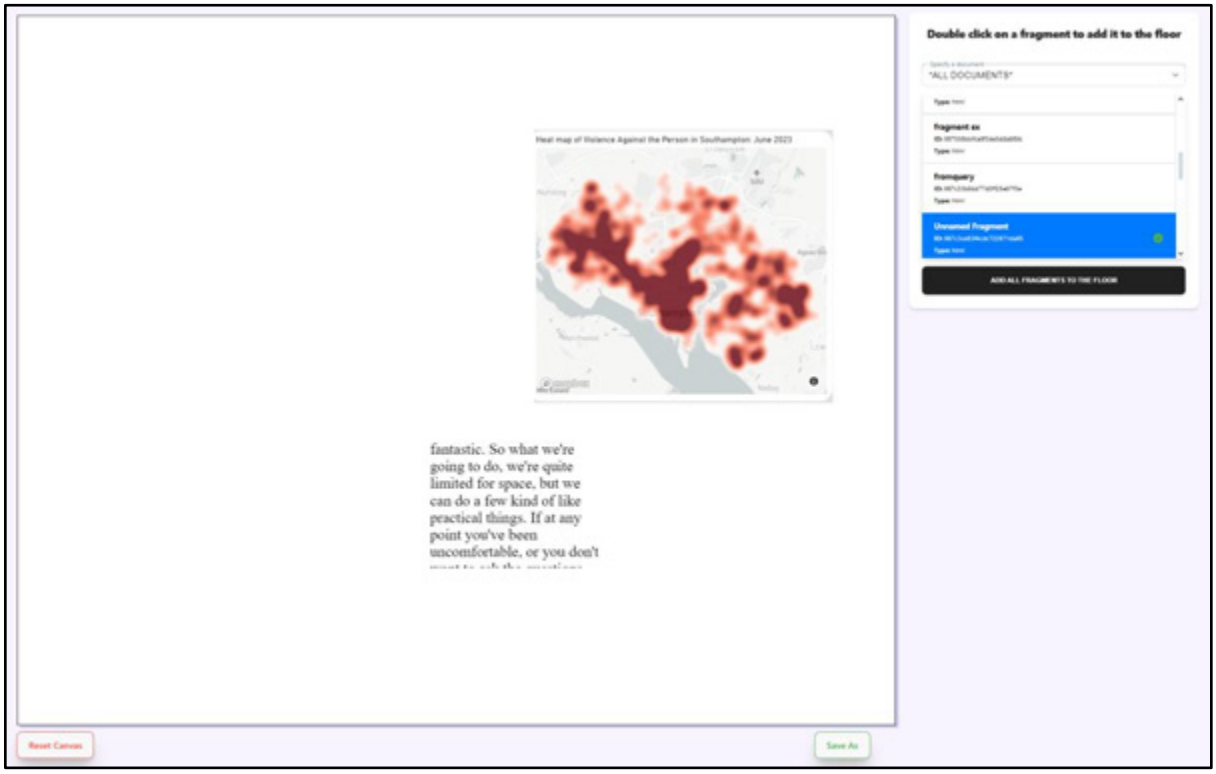
Screenshot showing workspace options and the menu to load existing virtual floors.

Clicking one of the open tabs in the workspaces section will quickly switch the view to that virtual floor (*and eventually document view).* When hovering over the button there is a small

icon to remove that tab from the open workspaces. The open tabs should persist locally when the site is reloaded.

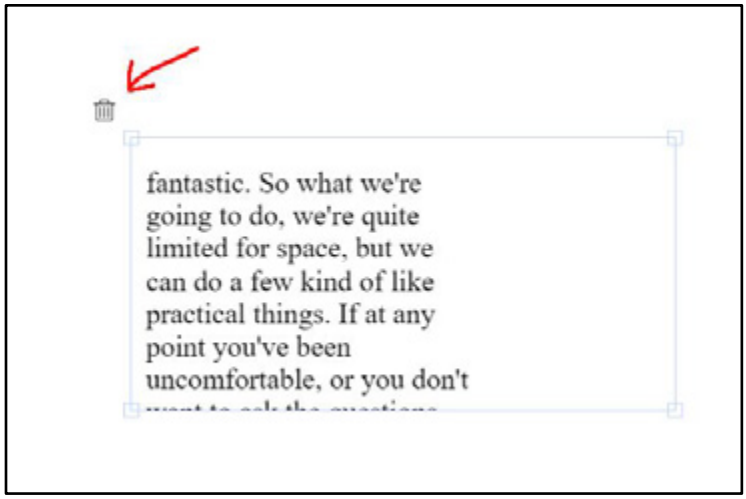
Screenshot showing use of the Sidebar when multiple workspaces are open.

When using a virtual floor, the user can select a number of fragments that they have extracted using a scrollable window at the top-right. This can be filtered by the document they were extracted from, in the same fashion as the **My Fragments** tool. Double clicking on a fragment will place a rendered object on the canvas, which represents the fragment. These can be resized and moved around the canvas at will. There are options to reset and save the virtual floor at the bottom of the screen.



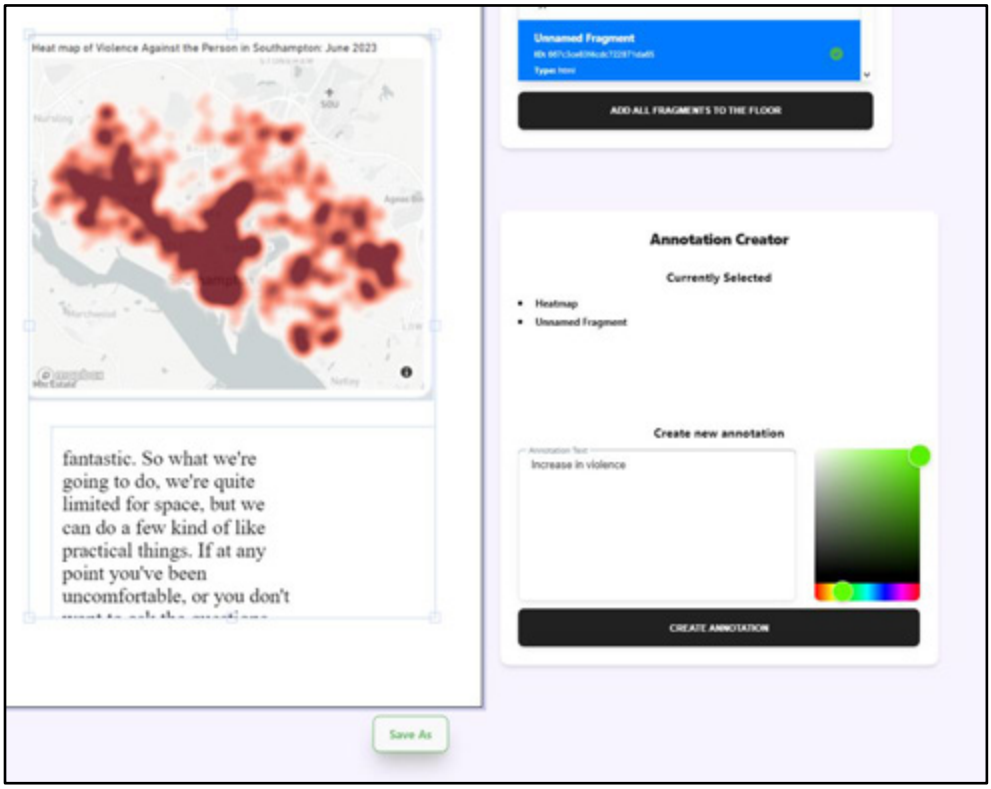
Screenshot showing basic Virtual Floor functionality.

A delete icon will hover near the top left of a fragment in the canvas when the fragment has been selected (i.e. clicked on), allowing the user to remove it from the canvas.



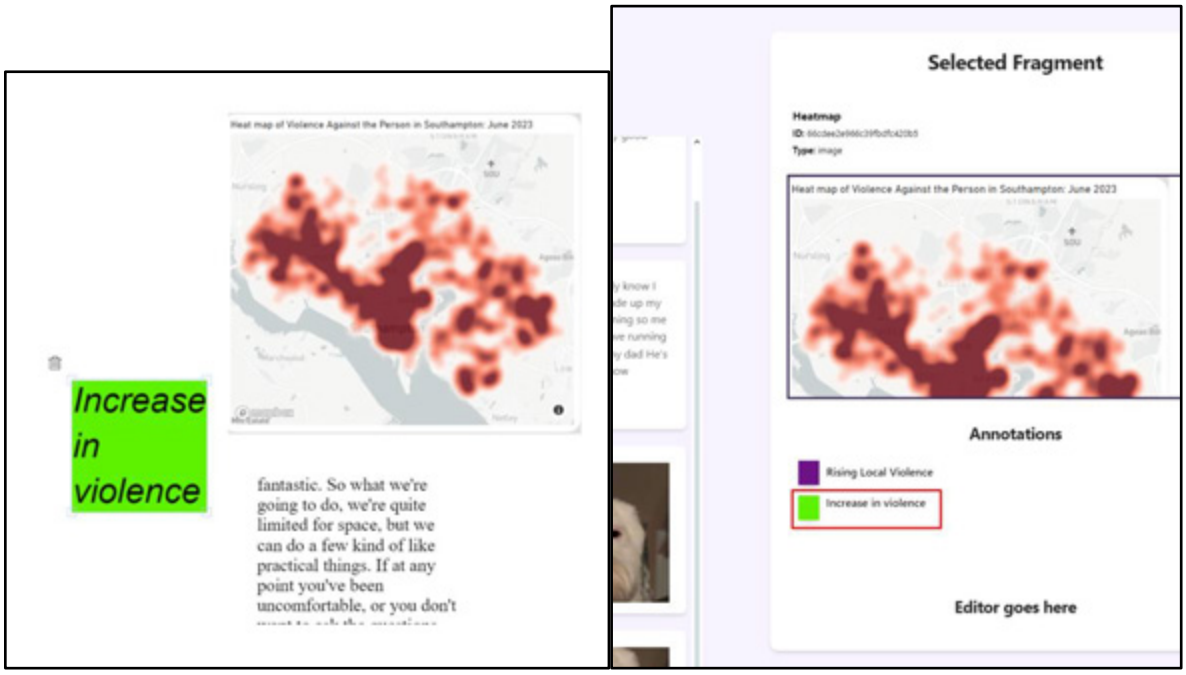
Screenshot showing the delete icon for fragments on the Virtual Floor.

When selecting fragments, both individually, and in groups, a window appears at the bottom-right for creating annotations. These annotations come in the form of a text string, and can be given a colour. This colour helps separate ideas and can be seen as a form of coding that is done in traditional qualitative analysis.



*Screenshot showing the **Annotation Creator** in use.*

Once the annotation has been created, a dialog box will appear on the virtual floor representing the annotation. This can be moved and resized in the same way as normal fragments. The annotation will now appear in the **My Fragments** tool, when relevant fragments are selected.



Screenshots showing the role of annotations in the system.

When saving a virtual floor, the user is given a prompt to enter a name. Upon re-saving, a separate save icon will appear for resaving the floor. Saving a floor/loading a floor from a save file will change the tab name from the default name to the new given name.

Screenshots showing the save functionality.

Technical Details

This is intended to serve the future developer of the NIE. The dataflow throughout the system is relatively complex, and due to restrictions of the React framework, is less modular than intended. Using this guide, future progress should be possible after a short while of acclimatising to the system for a new developer.

Overview of Frameworks & Important Details

- **npm** was used to manage the packages in the project.
- **Vite** is used to build the project, and was used because of its useful features like hot-reloading. Instructions on how to run the project exist on the Github page for the NIE: https://github.com/owenrichards02/NIE-prototype .
- **(*VM stuff*)**
- **Jotai** is a state manager used to handle consistency of various important states between components.
    - The atoms used are listed throughout this report. They are all defined in the ***state.js*** file. Some are constructed with atomWithStorage rather than atom, meaning they will persist in localStorage, rather than reverting to the initial values on a reload. This allows open tab information to persist when switching to and from tools, and reloading the site.
    - Fragments, documents etc. are stored in an atom and loaded from storage on the first load of the app. They are updated to and accessed from this atom, as well as recording changes in mongoDB. This limits API calls in certain scenarios.
    - **Jotai Dev Tools** is a package I installed to show the current value of atoms during development mode execution. The Jotai icon at the bottom left can be clicked on to see these details
- **MongoDB Realm** is used to store all user documents, fragments, annotations and saved virtual floors.
    - A cluster called NIE-cluster was created in the NIE Dev project, under the UoS Wellth Lab Organisation.
    - The cluster contains databases with collections for each type of stored data.
    - The **realm-web** npm package contains functions that can access and modify these collections as needed.
- **Tailwind CSS** is used throughout the application, in combination with traditional css classes in *App.css.* **Material Tailwind React** components have been used to improve styling, including the **Sidebar** and various cards in the fragment extractors.
- **Fabric.js** powers the canvas for the virtual floor. The **fabricjs-react** package is a version that is compatible with React.js. Some examples are on the npm listing: https://www.npmjs.com/package/fabricjs-react
- The file **icons.js** contains the dataURI source for the hovering delete icon, since it must be placed on the canvas, and therefore loaded as an Image from a dataURI.

NIE-Realm API

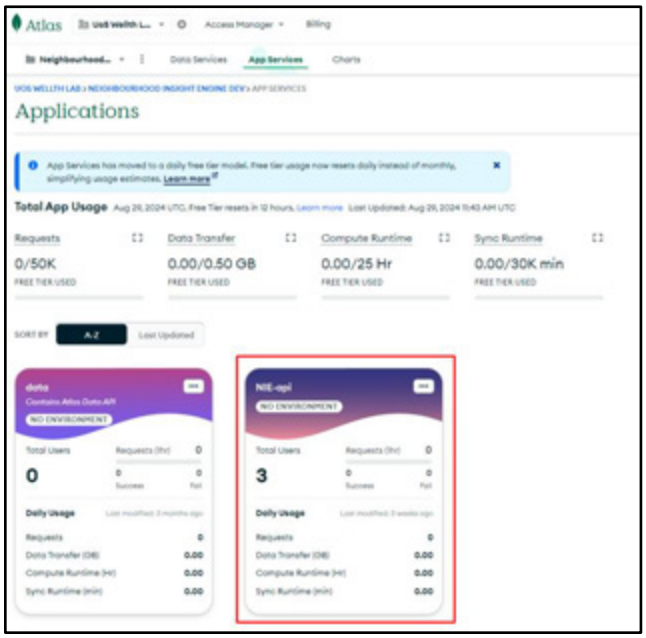
The NIE-Realm API is a collection of helper functions used for performing CRUD functions on the MongoDB collections that store documents, fragments etc. For instance, to find all documents that are stored, the *documents_findAll()* function can be called.

The majority of these functions are asynchronous, returning promises. They must therefore be used with the keyword *await*. These are often called from inside a UseEffect() React Hook, to support the use of results from asynchronous calls when updating states/atoms inside the NIE.

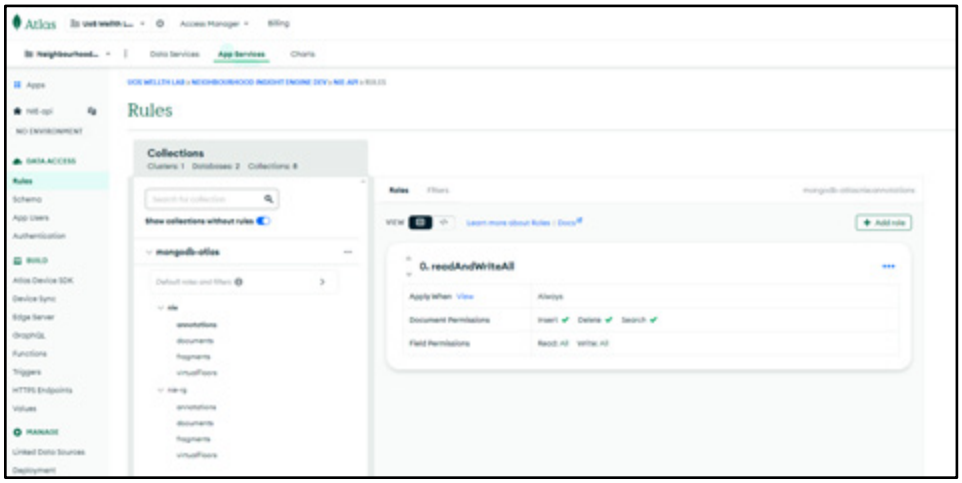
JSDoc comments provide more insight into how each function works. The index.html for the JSDoc can be found in the jsdoc folder in the NIE repository.

MongoDB + Connection

When creating new collections in mongoDB, you must configure the *Rules* for that collection in the NIE-api app. To ensure this, go to App Services on the cloud portal, and select the NIE-api app.



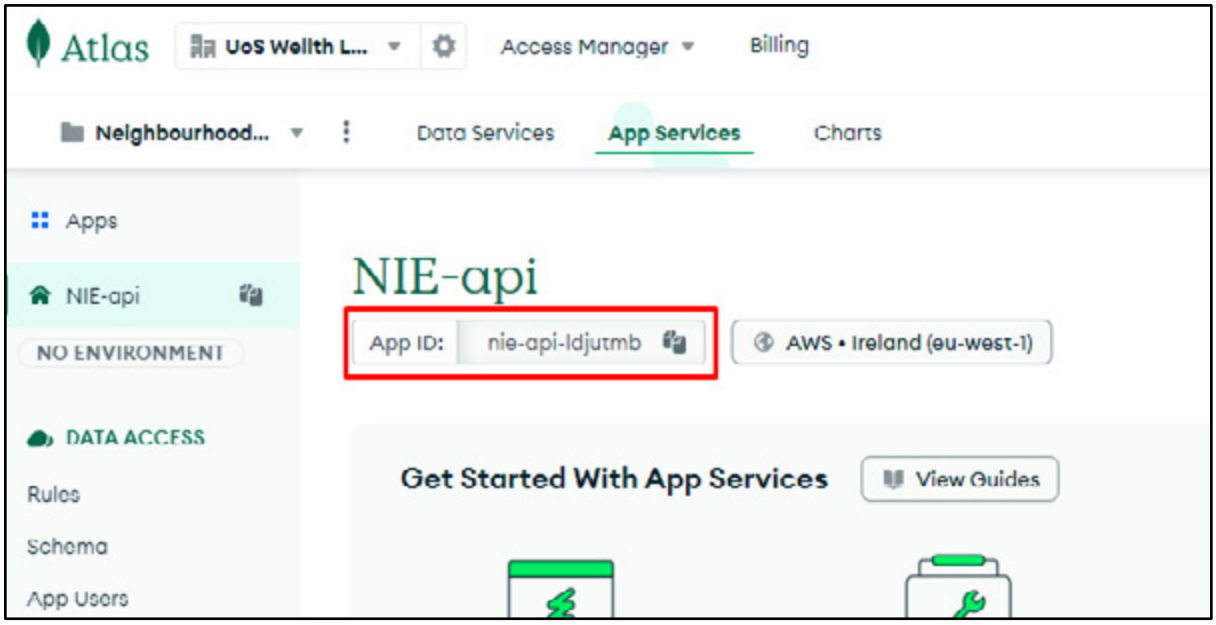
Then, add a Role for your collection, which is basically a permission tier that allows the access you need. Multiple rules can be added to create separate permission levels.



The contents of your .env needs to have the following:

```
VITE_REALM_APP_KEY='nie-api-ldjutmb'
VITE_REALM_EMAIL='or1g20@soton.ac.uk'
VITE_REALM_PASS='realmpass'
VITE_REALM_DB='nie'
```

The app key can be found here:



Storage Schema

```
Document => {ObjectID : _id,
             String   : name,
             Binary   : data,
             String   : html,
             String   : type,
             [String] : tags
            }

Fragment => {ObjectID : _id,
             String   : name,
             ObjectID : docid,
             Binary   : data,
             String   : html,
             String   : type,
             [[Int]]  : coords,   (for sections of images)
             [String] : tags
            }

Annotation => {ObjectID   : _id,
               String     : name,
               [ObjectID] : linkedFragments,
               String     : content, (in html format?)
               [String]   : tags
              }

Virtual Floor => {ObjectID   : _id,
                  Object     : floor,
                  String     : name,
                  }
```

floor is an object containing an a2c and an f2c object.

Navigation & Sidebar

- In App.jsx, a BrowserRouter is defined, providing the various routes for the tools in the app.
- Each tool is provided with a different route and buttons in the Sidebar use *useNavigate(route)* to change the view accordingly.
- Dialogs in the Sidebar are used to open/close new workspaces. They control the pop-up windows for loading existing/creating new workspaces.

Document Viewer

- The file uploader is set to accept .html files only.
- Uses a custom ItemList component that shows the ID and type of the document objects that are passed to it.
- When a new document is uploaded, the Jotai atom for documents is updated with the new document, as well as being uploaded to the collection in mongoDB.

My Fragments

- Fragment grid spaces out fragment grid icons incorrectly when there are 1<x<4 rows of fragments.

Fragment Extractors

- When a new fragment is created, the Jotai atom for fragments is updated with the new fragment, as well as being uploaded to the collection in mongoDB.

Create Image Fragment

- Create Image Fragment has a button to use the whole image as a fragment, but this functionality has not yet been implemented.
- Uses a library called react-image-crop for the cropping tool and to create the cropped image preview.
- It uses a modified version of the fragment creator called FragmentCreatorNoPrev, since the cropped image preview supersedes the role of the normal fragment preview.

Workspaces & WorkspaceArea

- The *tabs* atom keeps track of all open tabs.
    - It is a list of tab objects that look like this:

```
const newTab = {
        type: "floor",
        key: uuid,
        existing: false,
        index: tabs.length,
        name: "Virtual Floor " + tabs.length,
```

```
        vf_id: null
      }
```

- f2c and a2c are two atoms that keep track of the location and size of all fragments and annotations respectively in each canvas. E.g. At f2c[2], you can find the fragment information for tab index 2. They will be explained in further detail later.
- The Sidebar contains the Dialog/buttons which trigger the opening and closing of tabs by modifying the *tabs* and *currentTab* atoms.
    - Logic inside the *closeTab, chooseSavedFloor*, and *newFloorCreated* functions add a tab to the tabs atom and change the currentTab.
- The large UseEffect hook inside WorkspaceArea will be called on an update to the currentTab. Based on the currentTabsCount atom, it will find whether a tab has been created, removed, or whether a different tab has been switched to.
    - If a tab has been removed, sometimes the indexes of the f2c and a2c objects need to be changed.
- isVfReady is often set to false during updates, to both hide the canvas during the update, and to initiate a reload of the newly correct tab once it is set back to true, since a new VirtualFloor instance is created.

Due to some unfixed bugs, sometimes the app will not load, or crash every time you try to close specific tabs. To fix this temporarily, you can uncomment this section in WorkspaceArea.jsx, which will reset the atoms that persist in localStorage, allowing a full reset. Make sure to re-comment them before continuing to use the app:

```
    /* setTabs(RESET)
      setOpenTabsCount(RESET)
      setf2c(RESET)
      seta2c(RESET)
      setCurrentTab(RESET) */
```

Virtual Floors

f2c and a2c

- Based on the current tab index, references thisf2cRef and thisa2cRef are set up to reference the specific index of the f2c and a2c that represent the current tab.
    - Use thisa2cRef.current and thisf2cRef.current to access the values.
- Setter methods are created for this specific tab called setf2c_ForThisTab() and seta2c_ForThisTab(). Use this to update the values, as it avoids changing data for other tabs unexpectedly.

An example object inside f2c[tab_index] looks like this:

```
        const f2c_example = {
            frag: frag,
            locationObj: {
```

```
                    height: oImg.getScaledHeight(),
                    width: oImg.getScaledWidth(),
                    posx: posx,
                    posy: posy
                },
                canvasObj: oImg,
                uuid: uuid
            }
```

frag is the fragment object that can be found in the fragments atom or the fragments collection in mongoDB
locationObj is a custom 'Location Object' that represents an object's size and location.
canvasObj is the actual object generated when adding the fragment to the canvas.

Uuid's are generated using the crypto package, to help distinguish multiple copies of the same fragment, and also to match canvas objects to fragments in various cases, such as when a fragment is moved around the canvas. Canvas objects are given the uuid as an id field:

canvasObj.id == f2c_example.uuid
{*given they are referencing the same item on the floor.*}

Objects inside a2c[tab_index] are similar in nature, with some extra details that are hopefully self explanatory:

```
const a2lObj = {
                locationObj: locObj,
                canvasObj: annot,
                uuid: uuid,
                fragids: a2l.fragids,
                annot_id: a2l.annot_id,
                text: a2l.text,
                color: a2l.color
            }
```

fragids gives a list of all the *(mongodb NOT uuid)* ids of the linked fragments.

N.B. The majority of the time when mentioning the f2c and a2c, I am talking about the value for the current tab, not the entire list of all tabs.

Loading from localStorage

There are a large number of useEffect hooks at the top of the VirtualFloor function, each waiting for reloads to different states/atoms. Two states, *isReady* and *isAnnotsReady* must be set to true to initiate each part of the load.

- *isReady* is set to true when *fragmentList* is fetched from mongoDB and is not empty and f2c is also non-empty after being fetched from localStorage.
- isAnnotsReady is set to true when a2c is non-empty after being fetched from localStorage

The load() and loadAnnots() function will then place all the fragments/annotations on the canvas, and replace the old .canvasObj in each f2c/a2c object with the new canvasObj, as Fabric requires you to create new objects every time; it cannot load in existing ones. There is likely a way to combine the spawning from load and normal spawning functions, if more simplicity is required, or the app is becoming hard to maintain.

It is important to note that in a canvas with an empty a2c or f2c, the isReady and isAnnotsReady is set to true before the first NEW annotation or fragment is placed on the canvas, otherwise it will be reloaded and duplicated:

```
function spawnFragment(fragid){
        console.log("adding new fragment to canvas")
        if(thisf2cRef.current.length == 0){
            setLoaded(true) //avoids reload
        }
        ...
```

Canvas Setup

A large useEffect hook contains all the assignments of handlers for canvas events. To see how canvas events work, there is a demo on the Fabric.js site: http://fabricjs.com/events. This hook also contains statements to set the size, colour etc. of the canvas. These are also repeated in the *doReset()* function, as these changes are removed when performing editor.canvas.clear()

When a new canvas object is created, the rotation and stretching controls are disabled, for simplicity reasons. If you would like to re-enable this, you can remove the sections that look like the following:

```
    annot.controls = {
                ...fabric.Textbox.prototype.controls,
                mtr: new fabric.Control({ visible: false }),
                mt: new fabric.Control({ visible: false }),
                mb: new fabric.Control({ visible: false }),
                ml: new fabric.Control({ visible: false }),
                mr: new fabric.Control({ visible: false }),
            }
```

Fragment Rendering

Image fragments can be easily placed on the canvas, as they are loaded from the dataURI stored in the fragment object. Fabric has a function for creating an Image object from a dataURI, so this is simple:

```
if (frag.type == "image"){
    fabric.Image.fromURL(frag.html.split('\"')[1], function (oImg) {
                        doSpawn(oImg, frag, uuid)
                    }, {id: uuid});
```

N.B. in the example above, you can see how the uuid is passed to the new canvas object.

The easiest way we found to render all other HTML fragments was to convert the HTML to XHTML, so it follows XML rules, then create an SVG object with the XHTML as the body of a *foreignObject.* This can then be turned into a dataURI and used to create an Image on the canvas:

```
const doc = new DOMParser().parseFromString(frag.html, 'text/html');
            const xhtml = new XMLSerializer().serializeToString(doc);
            let svg = '<svg xmlns="http://www.w3.org/2000/svg"
display="block"><foreignObject width="200" height="200"><div
xmlns="http://www.w3.org/1999/xhtml">'
            svg += xhtml
            svg += '</div></foreignObject></svg>'

            fabric.Image.fromURL('data:image/svg+xml,' + encodeURIComponent(svg),
function (oImg) {
                doSpawnLoad(oImg, frag, locObj, uuid)
            }, {id: uuid});
```

The two main problems currently are that fonts *(i.e. css information)* are not preserved into the SVG, and the whole html section does not always fit inside the foreign object. This can be fixed, but I did not investigate it. I was following along from a site that claims to be able to fix both problems, https://ronvalstar.nl/render-html-to-an-image if these fixes seem necessary (there are JSFiddle sandboxes demonstrating the changes required).

Handling Modified Objects

The methodology behind this is to search through all objects in the *object:modified* event callback object, match their uuid to an object in the a2c or f2c, and update their *locationObj* in this f2c/a2c object. This ensures that upon reloading the tab, the new location & size are saved.

On a single object this works well, however when multiple objects are moved simultaneously *(i.e. select multiple and drag all)*, the new locations in the event callback object are relative to the new location of the selection; only the actual location of the whole selection are stored in absolute values. The original location of the selection is also stored in this event, so an x/y offset can be calculated, then added to the x and y positions of the original objects.

A temporary rolling f2c/a2c are used here to ensure the async React state doesn't lose track of any of the changes.

Keeping Track of Selections

Initially, the multi-select key is 'Shift', so to change it to the more traditional 'Ctrl' I set it here:

```
if(editor?.canvas){
        editor.canvas.selectionKey = "ctrlKey"
  }
```

The *selectedObjects* state (and reference) is used to keep track of what is selected at all times. I add just the canvas objects for simplicity reasons, as this is never saved, so uuids can be matched to source objects at runtime. Separate events are called for when a selection is created, changed (e.g. an extra item is added to the selection) and cleared.

The length of *event.selected* can indicate whether it is a multi select. *Event.deselected* is also useful to help remove those that have been deselected from our list. The *selectedObjects* state is then used for things like the annotation creator.

A delete icon will be placed on the canvas when a single object is selected; deleting multiple is currently not allowed. The selection handlers often remove and add the delete icon depending on the current number of selected objects.

A long *removeFromStorage* function is called when the delete button is clicked, as well as removing the button from the canvas. This function will match the uuids of the selected object by referencing the selectedObjects state, and remove it from a2c/f2c.

If an annotation is removed, it is also removed from the annotations collection in mongoDB.

Annotation Creator

The list of selected fragments is created by referencing the uuids of selected canvas objects with the fragments atom. Each matching fragment has its .name value placed in *fragNamesList,* which is then displayed to the user.

The background colour picker is from a module called react-colourful.

The annotation creator will update the annotations atom directly, before calling the onAnnotCreated method in the Virtual Floor. Annotations on the canvas are simple Fabric textboxes, that are given a single colour and the text from the annotation creator window. This method will create it and place it on the canvas near the top left of the selection, however it may be more effective to place it in the nearest open space if possible.

Loading & Saving Floors

The floor is saved with a name and a ‘floor’ object, containing both the f2c and a2c (for this tab!). When loading from a save, the values in this object are used to overwrite the local f2c and a2c, which will then have the same effect as loading from localStorage.

To achieve this load, when creating a new tab, the Sidebar will set *tab.existing* to true, and *tab.vf_id* to the mongoDB _id for the saved virtual floor. This allows WorkspaceArea to recognise that it needs to load the new f2c/a2c for this tab from one of the objects in the virtual floors atom.

The code for the save dialog can be found at the bottom of VirtualFloor.jsx, which is where the pop-up window for saving a floor is defined.

When saving floors to mongoDB, sometimes the structure of the canvasObject would have a cyclic link, such that the JSON parser in the realm-web library couldn’t handle the f2c/a2c objects. To handle this, the ‘stringify’ and ‘parse’ from *flatted* can be used, which is a library that deals with parsing cyclic links in JSON.

Miscellaneous/Considerations

- Document/fragment HTML is stored as UTF-8 text. If >16MB, GridFS must be used, but this is currently not implemented.
- All images uploaded have been converted to data URIs and then uploaded with the ‘image’ type.

Example Excel Survey HTML Format

```
<div data-surveyID='x'>
      <ul data-responseID='x'>
            <li class="qna">
                  <p class="question"><b>questionText</b><\p>
                  <p class="answer">answerText<\p>
            </li>

      </ul>
</div>
```

Example Transcript HTML Conversion

```
<div data-interviewID='x'>
      <ul>
            <li class="dialogue">
                  <p class="speaker"><b>speakerName</b>: time<\p>
                  <p class="speech">speech<\p>
            </li>

      </ul>
</div>
```

Future Goals

Many of these future goals have evolved from discussion with those involved in the project and intend to provide new use cases or improvements to existing use cases for the NIE. A list of known issues/bugs is given to help future developers iterate on the existing prototype.

Necessary Features/Changes

- Deletion functionality for old fragments, documents, annotations and saved virtual floors.
    - Some API functions are available for this already.
- Confirmation of successful/unsuccessful saving and re-saving of virtual floors.
- Use percentage positioning of components rather than pixels, to support multiple resolutions. The app currently looks very different on 1440p vs 1080p.

Desired Features

- Zoom/scrollable virtual floor - pan around in an infinite canvas.
- Document viewer displays all annotations (similar to MS Word comments)
- Multi user support.
- Interactive document view: fragment extraction by highlighting sections.
- Other annotation types e.g. relations between groups.
- Hover over fragments on the virtual floor to see linked annotations and details.
- Improved visuals, styling consistency, UI etc.

Known Issues

- The button which adds all available fragments to a floor does not update the f2c and a2c objects correctly, causing them to not exist upon reloading the floor.
- Fragments not scaling properly inside SVG render, not preserving css. An explanation is in the fragment rendering section.
- Occasionally re-loading from localStorage after closing tabs will break the app. Unsure of a cause.

Find out more:
www.andtowns.co.uk/project/neighbouring-data

This report should be cited as:

Daniel Ashton, Richard Gomer, Joey Jones, Nicky Marsh, Joseph Owen, and Owen Richards (2025). *Transforming the Observatory: From Archive to Engine*. University of Southampton.
DOI: **10.5258/SOTON/P1223.**

Photo credits:

Page 8, 9: Daniel Ashton
Page 3, 4, 5, 6, 10, 13: Ash Sealey