WELCOME TO COLOURING BEIRUT DATA ETHICS POLICY

INTRODUCTION

Trust in Colouring Beirut, and adherence to high data ethics standards are viewed as crucial to the platform's success and to its longevity. This may relate to trust that stated aims and objectives of the platform are actually as described and that contributions will be used for the common good; that data are accurate; that building occupants' privacy is being respected, as well as the privacy of platform users; that the principles of partners organisations are upheld, and/or that the long-term management of the project is secure. Breakdown of trust in any of these areas is considered to pose a major risk to the sustainability of the project.

Respect for each user and recognition of the importance of each contribution is also central to the project's approach. Colouring Beirut is designed for everyone. It looks to create a safe, positive, constructive space for users of all genders, ages, cultural backgrounds and abilities to enjoy and benefit from, and a place where the sharing of knowledge by the user is both appreciated and respected.

The Open Data Institute's Data Ethics Canvas is used by Colouring Beirut to help identify and address potential ethical concerns.

THE DATA ETHICS CANVAS

Data ethics is described by the Open Data Institute (ODI) as a

" *A branch of ethics that evaluates data practices with the potential to adversely impact on people and society-in data collection, sharing and use.*"

Ethical use of data brings about trust and helps allow data to work for everyone.

Colouring Beirut is using the ODI Data Ethics Canvas to help identify and

manage ethical issues throughout the lifecycle of the project.

As part of the process of development, existing, and new features within the platform are checked against the questions posed by the ODI Ethics Canvas. First stage responses to core questions are given below.

WHERE ARE DATA FROM? ARE PERSONAL OR SENSITIVE DATA INVOLVED?

Most of the datasets we collect and make accessible relate to the physical characteristics of buildings which are able to be viewed by the public already from the street or from a city image. These include information on a building's use, age and type, whether it has just been demolished and its green context. Open data sets included within the platform, relating to regulation and performance, such as a government building's energy rating or whether it is it listed or in a conservation area, will also already be within the public domain. Our job is imply to make them more accessible and to help increase their accuracy through verification. For data relating to the construction and demolition history of buildings , information will be derived from historical texts maps and archive records. Information on construction teams may involve personal knowledge, public data held by industry, but in most cases will be derived from historical records. For the Like me? section the users personal view on whether the building contributes to city is requested.

Our data are gathered in three ways. Firstly by identifying large-scale datasets held by central and regional government bodies, and large publicly funded organisations, able to be released as open data. Secondly from crowdsourced submissions from multiple sources (ranging from from residents, community-led planning organisations, schools, research institutions, housing providers, local authorities, commercial firms). Thirdly, through spatial analysis and the use of computational approaches (in relation to physical attributes only), with the latter allowing existing attribute information to be combined to predict and prepopulate new attribute data for users to verify.

We do not collect personal data, and actively discourage users on our 'sign up' page from contributing email address or real names to the site. The sign up agreement also emphasises that when users make a contribution to Colouring Beirut, they are creating a permanent, public record of all data they have added, removed, or changed; that the database records the username and ID of the user making the edit, along with the time and date of the change, and that all of this information is also made publicly available through the website and through bulk downloads of the edit history.

Additional checks have also been put in place to ensure privacy on the part of building occupants, owners and users. The site uses a dropdown menu system for all data submissions and does not allow free text. An internal moderation system for data sources is also being developed. Colouring Beirut

does not collect data about a building beyond its external walls, other than with regard to its land use and the number of self-contained units within it. With regard to building ownership data, are only collected on state and 3rd sector assets. The Like me? Category does not include a dislike button to discourage casual malicious behaviour, specifically with regard to the domestic stock. Here the category is not included to collect scientific data but instead to encourage debate and discussion amongst citizens around the sustainability of the stock.

HOW ARE WE ADDRESSING ACCURACY, BIAS AND INCOMPLETENESS?

To help address issues of accuracy and bias a number of features have been included, with each subcategory having an accessible edit history, a source box, a verification button, and a query button to enable problems that cannot be addressed within the editing system to be raised. Moderated dropdown options for sources range for 'non-expert assessment viewed at first hand' or 'from an image', or ' expert assessment viewed at first hand', to a reference to a specific 'scientific paper'. Edit histories to assess the knowledge base of previous editors. Phrasing of specific subcategory questions is also in certain cases used to address uncertainty, with data are collected at the most granular scale wherever possible.

The main editing page contains a clear statement that data are derived from multiple sources and that accuracy of the data must be determined by the user.

As with the Wikipedia model the project is being planned as a low cost model, overseen by expert contributors. Our stewarding structure is currently being developed.

WHO ARE WE SHARING DATA WITH AND UNDER WHAT CONDITIONS

 Colouring Beirut has been designed by University College Beirut as a free knowledge exchange platform that collates, collects and generates open data on Beirut's building stock, able to be used by everyone. Datasets can be accessed by viewing our online maps or by downloading from our site. As soon as permission is given by Ordnance Survey for UCL to release location/footprint geometry data we will. Until then all data will released as open data tables, with spatial references included where possible - i.e. where OpenStreetMap coordinates have been able to be linked, or other open location data integrated.

The site is explicit on the user agreement required on our sign up page on the way that contributed data can be used, and that we cannot accept restricted data.

Colouring Beirut contributions are open data, licensed under the Open Data Commons Open Database License (ODbL) by Colouring Beirut contributors.

Users are free to copy, distribute, transmit and adapt our data, as long as they credit Colouring Beirut and its contributors. If users alter or build on Colouring Beirut data, they may distribute the result only under the same license

They are free to copy, distribute, transmit and adapt the data, as long as they credit Colouring Beirut and our contributors. If they alter or build upon our data, they may distribute the result only under the same license.

Colouring Beirut emphasizes local and community knowledge. Contributors use a variety of sources and local and expert knowledge of buildings to contribute data and verify that Colouring Beirut is accurate and up to date.

When users contribute to Colouring Beirut, they make their contributions available as open data for anyone to copy, distribute, transmit and adapt in line with the license.

UCL is unable to accept any data derived from copyright or restricted sources, other than as covered by fair use. Full attribution of data sources where appropriate.

The sign up agreement emphasizes that when you make a contribution to Colouring Beirut, you are creating a permanent, public record of all data added, removed, or changed by you. The database records the username and ID of the user making the edit, along with the time and date of the change. All of this information is also made publicly available through the website and through bulk downloads of the edit history.

All are data are free to use subject to the terms of a liberal open data license (ODbL). We do not sell data and we will not share user's personal data (such as their email address) with other organisations.

Our platform code are also open and we encourage its use by other cities and towns. Code are available on our GitHub site

https://github.com/tomalrussell/colouring-Beirut' under the following licensing terms:

'Colouring Beirut Copyright (C) 2018 Tom Russell and Colouring Beirut contributors'. This program is free software: you can redistribute it and/or

modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version. This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.

See the GNU General Public License for more details:

WHAT RIGHTS WILL THE SOURCE HAVE?

As noted above, under the Open Data Commons Open Database License (ODbL) Colouring Beirut contributors are free to copy, distribute, transmit and adapt our data, as long as they credit Colouring Beirut and its contributors. If users alter or build on Colouring Beirut data, they may distribute the result only under the same licence

They are free to copy, distribute, transmit and adapt the data, as long as they credit Colouring Beirut and our contributors. If they alter or build upon our data, they may distribute the result only under the same licence.

ARE WE SURE WE ARE NOT CONTRAVENING ETHICAL FRAMEWORKS

Trust in the platform is fundamental to its success. Data privacy and data ethics are of the highest priority to the project. We actively discourage the contribution of personal data, avoid the collection of data within a building's fabric, incorporate controlled dropdown menus with an internal moderation system for sources. Our close working partnerships and discussions threads provide ways in which to capture comments regarding ethical issues and we are actively seeking to learn from, and collaborate with organisations advancing the data ethics agenda.

WHY ARE WE COLLECTING DATA? ARE WE REPLACING A SERVICE? ARE WE MAKING THINGS BETTER AND FOR WHOM?

We are collecting and collating open data on Beirut's building stock to provide free access to essential information for citizens, researchers, education providers and policy makers. Also to provide information for those designing, constructing, caring for and managing Beirut's buildings.

Our aim is to provide the first port of call for open data on Beirut's building stock. The release of data is also designed to stimulate the production of innovative and efficient products within the academic, non-profit and commercial sectors

We also believe that it is healthy for platforms collating data on the building stock to be curated by universities, and others, whose stance is impartial, and whose brief is to undertake research for the public good.

ARE WE CLEAR IN THE WAY THE DATA WILL BE USED?

We are already aware of research areas, such as energy analysis, building lifespan analysis and stock resilience, where demand for building attribute data is very high. We also know , from extensive consultation, that demand for building level data at city scale exists from construction and property industry, housing suppliers, planning bodies and the education sector. We are therefore excited about the many ways the data may be used.

We are currently developing a data showcase facility to allow users to upload ways in which the building attribute data from Colouring Beirut, and from other sources around the world are being used to solve urban problems, and to in doing so to inspire and inform.

WHO WILL BE POSITIVELY IMPACTED AND HOW? HOW CAN WE MAXIMISE AND MEASURE THIS?

As noted above, access to building attribute data for Beirut at building level will impact on those involved in the design, research, construction, management and maintenance and analysis of Beirut's buildings. It is also designed to provide a resource for use in education and to encourage interest from Beiruters as a whole to make the city more sustainable. Our project is also designed to encourage use by diverse audiences - in terms of gender, cultural background, ability and age. This element is central to the project's design

Our task at the moment is to begin to release data on all our 50 subcategories. The second stage of the project will involve the introduction of features relating to the monitoring of platform use (using analytical software), the analysis and application of the data.

WHO COULD BE NEGATIVELY AFFECTED BY THE PROJECT? & HOW IS tHSI BEING ADDRESSED?

The ODI's 10th item on its Data Ethics Canvas addresses the issue of negative project impact? Could the manner in which this data is collected, shared and used cause harm? Or be used to target, profile or prejudice people, unfairly restrict access? Could people perceive it as harmful?

As noted above the free availability of satellite and streetview imagery city streets and buildings have now been able to be interrogated for many years. 3D satellite imagery models are now also available for

much of Beirut. Most Colouring Beirut categories relates to the physical characteristics of buildings, already able to be viewed on these models, or in person from the street. Datasets such as whether a building is protected and its energy rating are already publicly available. Property websites also already contain large amounts data on citizens' homes including interior images. We do not collect images or data on building interiors.

Owing to concerns raised with regard to privacy and ownership data Colouring Beirut only collects data on buildings where the freehold is held by the state or 3rd sector owners, with private ownership included as a default colour. The 'Like me?' category is also being closely monitored.

All spatial data projects that collect information able to be linked to specific locations need to be responsible with regard to the type of data collected and the way it is collected. The main ways in which we are working to minimize negative impacts are by a) not collecting personal data, b) not collecting data on the insides of homes, c) using preset dropdowns and moderated ' sources', and d) setting up and ethics advisory group and an alert system to highlight existing potential future issues e) allowing one vote per user on like me and having no negative option, e) having a sign in page that provides clear guidelines for responsible and ethical use of the site.

HOW CAN PEOPLE APPEAL OF AFFECT CHANGES TO THE SERVICE?

Users will be able to send us feedback on site improvement forms and flag if they are concerned about an entry. These features are currently being developed. Comments can also be made on existing discussion threads.

HOW ARE WE BUILDING IN CONSIDERATIONS OF PEOPLE AFFECTED BY OUR PROJECT?

The ODI's 13th Ethics Canvas item addresses considerations of people affected by our project. Are we creating potential risks or issues? How are limitations being communicated to those the data is about, and those impacted by its use? And how are we doing this?

Colouring Beirut has been designed from the outset in consultation with representatives from diverse sectors. These are listed on our 'Who's involved page'. The platform is being developed as an interdisciplinary knowledge exchange project. Our aim is to work with our partners and their networks to allow possible risks, concerns and improvements to be raised at the earliest possible stage - through our public discussions forum and feedback forms. For issues to then be addressed and features added, adjusted and/or removed as appropriate.