TOWARDS A NATIONAL COLLECTION

UKRI Arts and Humanities Research Council

FIRST REPORT
DISCOVERY PROJECTS

Transforming Collections:
Reimagining Art, Nation and Heritage

OCTOBER 2022

University of the Arts London | Tate

Table of Contents

Authors

susan pui san lok (UAL); Mick Grierson, Anjalie Dalal-Clayton, Athanasios Velios (UAL); Christopher Griffin, Hannah Barton (Tate); Ananda Rutherford, Charlotte Webb, Kit Bower-Morris (UAL); Jerneja Rebernak (UAL), Fleur Kaminska (UAL)

Executive Summary

Transforming Collections is led by University of the Arts London (UAL) Decolonising Arts Institute in collaboration with UAL Creative Computing Institute, and close partnership with Tate as an Independent Research Organisation. Our additional 15 project partners are: Arts Council Collection, Art Fund, Art UK, Birmingham Museums Trust, British Council Collection, Contemporary Art Society, Glasgow Museums, iniva (Institute of International Visual Art), JISC Archives Hub, Manchester Art Gallery, Middlesbrough Institute of Modern Art (MIMA), National Museums Liverpool, Wellcome Collection, the Van Abbemuseum, Eindhoven, and the Government Art Collection.

Transforming Collections is underpinned by the belief that a national collection cannot be imagined without addressing structural inequalities, engaging debates around contested heritage, and revealing contentious histories embedded in objects. In 1999, the late sociologist and cultural theorist, Stuart Hall, posed the question 'Whose heritage?'. Hall called for the 'unsettling' and 'reimagining' of heritage and nation.[1] More than 20 years on, the need to critically question and transform notions of 'heritage' and 'nation' remain as urgent as ever.

The project combines critical art historical and museological research with participatory machine learning (ML) design. A series of artistic research residencies will embed creative activations of ML within the project's research and public programme. The project seeks to surface suppressed histories, amplify marginalised voices, and re-evaluate artists and artworks long ignored or side-lined by dominant narratives and institutional practices. It aims to enable search across collections, to surface patterns of bias and relations of power, and interrogate policies and practices of classification, categorisation, description, and display. Critically and creatively connecting and disrupting collections, this interdisciplinary collaboration will open up new interpretative frames to imagine an evolving 'national collection', at once connected yet distributed. *Transforming Collections* aims to 'unsettle' and enrich existing knowledge with multiple and multivocal narratives, and 'potential histories' of art, nation and heritage.[2]

The project is designed around five parallel and interweaving work strands:

- Strand A: Surfacing Bias Across Collections (critical case studies)
- Strand B: Resurfacing Artists and Artworks Across Collections (critical case studies)
- Strand C: Participatory Design of ML (ideation and refinement workshops)
- Strand D: Interactive ML Technology Development (prototyping and testing)
- Strand E: Public Engagement Programme (in person and online events)

Collections case studies and ML development are progressing in tandem, with regular cross-strands meetings to ensure that critical research and design processes are mutually informed and driven. Participatory design and feedback mechanisms are being established across the partnership to embed and sustain collaborative ethical approaches. As the project unfolds, invited workshop participants will include diverse academic researchers, staff and students, as well as representatives from concurrent Towards a

1 Keynote speech by Stuart Hall given on November 1, 1999, at the National Conference, Whose Heritage? The Impact of Cultural Diversity on Britain's Living Heritage held in Manchester, England.
2 Ariella Aïsha Azoulay, Potential History: Unlearning Imperialism (London; Brooklyn, NY: Verso, 2019).

National Collection (TaNC) programme projects. Proposals for cross-TaNC projects activity on data ethics are also in development.

The first year of the *Transforming Collections* project has focused on extending and expanding the collections audit data from the Black Artists and Modernism project (AHRC 2015-18); auditing Tate Collection's Subject Index tags; preparing the digitisation of the unique artists' archive within the Stuart Hall Library at iniva; developing collaborative interdisciplinary working processes; developing data sharing guidance and ethical working principles; undertaking initial partner workshops; identifying early case studies; gathering diverse datasets from Tate and our various partners and organisations; and initial public programme development including a call for artists in residence. *Transforming Collections* will culminate in a public programme in autumn 2024.

Abstract

Transforming Collections aims to enable digital search across collections, to uncover patterns of bias in collections systems and narratives, to reveal hidden connections, and to open up new interpretative frames and 'potential histories' of art, nation and heritage.[3]

Whose voices, bodies and experiences are centred and privileged in collections? This project is underpinned by the belief that a national collection cannot be imagined without addressing structural inequalities, contested heritage and contentious histories embedded in objects. In 1999, the late sociologist and cultural theorist Stuart Hall posed the question 'Whose heritage?'. Hall called for the 'unsettling' and 'reimagining' of heritage and nation. More than 20 years on, the need to critically question and transform notions of 'heritage' and 'nation' remain as urgent as ever.

Our approach brings together academic and artistic research into collections and museum practices with participatory machine learning (ML) design. A series of artistic residencies will lead to new works that critically and creatively activate the emerging research and ML tools. A major public programme with Tate Learning will generate insights and understandings of the ways in which the *Transforming Collections* research can enable new stories to be told.

Led by UAL in close partnership with Tate among our 16 partners across the UK, the project seeks to surface suppressed histories, amplify marginalised voices, and re-evaluate artists and artworks long ignored or side-lined by dominant narratives and institutional practices. We want to imagine an inclusive, evolving, (re)distributed 'national collection' that builds on and enriches existing knowledge with multiple and multivocal narratives, to critically connect and imaginatively disrupt collections, and transform them.

3 Azoulay, Potential History.

Aims and Objectives

Transforming Collections aims to dissolve barriers between collections by addressing digital search and research capability as entwined priorities, with public engagement conceived as integral to the research development, design, and dissemination. Bringing together decolonial intersectional feminist praxes and creative machine learning, the project's approach is driven by the following questions:

- Whose heritage? Whose voices, bodies and experiences are centred and privileged in collections? What are the problematic concepts used to label, classify, categorise and describe objects, and how do these perpetuate racist, xenophobic, misogynist, heteronormative, white supremacist and ableist views?
- How can we surface and transform the architectures, algorithms and relations of oppression that structure and define collections' narratives? Can decolonial feminist approaches to machine learning help to counter or resist the replication of colonial and imperial modes of perception embedded in cataloguing and interpretation practices? How can we ethically engage and empower diverse stakeholders in transforming the politics and processes of representation?
- What could an equitable, inclusive, distributed, connected, evolving 'national collection' look like? How might we look across collections to create multivocal narratives that enrich knowledge within collections? What inspiring, unexpected or uncomfortable stories could such a national collection tell?

The project builds on the insights and findings of previous research projects led by UAL, namely the AHRC Black Artists and Modernism project (BAM, 2015-18) and the UKRI MIMIC project (Musically Intelligent Machines Interacting Creatively, 2018-21), as well as the recent Tate-led TaNC Foundation Project, Provisional Semantics (2020-22).

The project's objectives are to develop, model and test critical, creative and ethically grounded and inclusive technological approaches that enable cross-search of collections and increase research capability by surfacing patterns of bias, revealing hidden or unexpected connections, and opening up new interpretative frames and potential narratives of art, nation and heritage.

Transforming Collections will:

- Recruit diverse researchers and practitioners of colour and ensure at least 50% representation in the project team to support research capability, capacity and early careers into the future
- Produce a series of critical case studies underpinned by decolonial and intersectional feminist approaches that inform – and are informed by – ML development. These will be disseminated through workshops/seminars, a conference, online/peer-reviewed publications, and a project website
- Develop adaptable and transferable supervised ML software through participatory design processes, with interactive ML methods. This will enable researchers, curators and wider audiences to cross-search collections and diverse datasets; uncover unexpected connections; surface hidden works and histories; and empower diverse users to control the process of determining significance
- Digitize Iniva's uncatalogued archive of c.5,000 slides and c.200 artists' files, to establish an important new dataset and vital resource within the Stuart Hall Library at UAL; make searchable

hitherto hard to access documents relating to artists of colour[4]; and provide a critical counterpoint to objects and data both present and absent in UK collections

- Build, host and maintain supervised ML software (algorithm and interface) at UAL for a minimum of 10 years on the CCI public GitHub, with resources to support dissemination to partners for beta testing and potential wider rollout
- Develop and deliver an engagement programme including an international conference, workshops/seminars, papers/essays towards an edited journal issue, a book publication, and an accessible project website for general audiences
- Develop and deliver a major culminating public engagement event as part of Tate Learning Projects and Programmes, including the physical installation and online display of artists commissions and an extensive programme of participatory workshops and talks
- Collaborate with Art UK and partners to showcase artists and artworks online

4 Transforming Collections refers to 'Black artists' for consistency with the term used in the audit initiated by the AHRC Black Artists & Modernism (BAM) project, significantly extended and expanded by the AHRC TaNC project. The project also refers to 'artists of colour' while recognising that artists may choose to identify by neither or different terms, as suggested by the #BAMEover statement for the UK.

Partnership Structure

Lead Institutions

University of the Arts London's (UAL) **Decolonising Arts Institute** (DeAI) is leading on the project's conceptual and strategic direction and delivery, including project management, planning, coordination, administration, and communications. DeAI is working in close collaboration with the **Creative Computing Institute** (CCI), whose team is leading on the participatory co-design and technological development strands of the project. As UAL's main partner and Independent Research Organisation, **Tate** co-leads, co-develops and co-delivers the project, contributing expertise, research, insights, facilitating access to Tate Collection data, and co-curating the project's public engagement programme.

Project Partners and Collaborating Organisations

Initially launching with 14 partners in addition to Tate, *Transforming Collections* now has a total of **15 partners** including the Government Art Collection who joined the project in July 2022. The *Transforming Collections* partnership reflects the nationwide scope and ambition of the project to engage, benefit and positively impact collections across the UK, large and small, and builds on UAL's long history of collaboration with Tate, as well as previous projects with 12 of our partners. These include the major AHRC Black Artists & Modernism project (2015-18) led by UAL in partnership with Middlesex University, and several DeAI projects: the Decolonising British Art seminar series (2020-21, supported by British Art Network and Paul Mellon Centre); the Decolonising Collections research residencies (2021-22, supported by Art Fund); the Digital Artist Residency (2021-22, in collaboration with iniva), and the workshop/seminar series, Doing the Work: (2021-22, in collaboration with Contemporary Arts Society).

Partner organisations across museum, collection, arts, charity and archive sectors will contribute collections management and curatorial research insights, support access to collections databases, participate in ML co-design workshops, as well as engagement and dissemination activities. The partnership includes four Collaborating Organisations who are direct beneficiaries of the project funding. Selected collections will also host and support artists of colour in digital residence starting from January 2023.

The **Arts Council Collection, British Council Collection, Birmingham Museums Trust, Glasgow Museums, Government Art Collection, Manchester Art Gallery, National Museums Liverpool, Middlesbrough Institute of Modern Art and Wellcome Collection** are facilitating access to their collections management data and records, engaging in curatorial support and participating in selected workshops/seminars. **JISC Archives Hub** will enable access to its numerous art collections among its 350 repositories.

The major arts charities, **Art Fund** and **Contemporary Art Society**, are providing data on their UK-wide collection acquisitions and donations since the 1900s, including grants awarded, gifts and bequests, artists and exhibitions profiled through the magazine *Art Quarterly*, and organizational histories.

Art UK is providing analytics on its aggregated collections data on c.48,000 artists, c.250,000 artworks, user-generated tags, descriptions, and artworks online.

iniva is digitising its unique artists' archive, which is physically housed within the Stuart Hall Library. iniva is an independent organisation located at UAL's Chelsea site.

Van Abbemuseum in the Netherlands will co-host an international conference in April 2023 to situate the project's key questions, emerging research and ML development, in relation to artistic and curatorial interventions into 'national' and 'international' collection praxes.

Staffing Structure

Project Management and Administration

The Principal Investigator (PI) and project team are supported by Project and Partnerships Manager **Jerneja Rebernak**, and Project Administrator **Fleur Kaminska**. The PI is further supported by a **Management Board** which meets every six months to track the project's progress against identified objectives and milestones. The Management Board members are: Benjamin Stopher (Dean of UAL CCI), Prof Malcolm Quinn and Prof Pratap Rughani (UAL Associate Deans of Research), Emily Pringle (Tate Head of Research) and Danielle Tran (UAL Director of Education). An external **Advisory Board** meets formally once a year and informally as needed, to contribute cross-sector expertise, and ensure the relevance and impact of the project. The Advisory Board members are: Melanie Keen (Director, Wellcome Collection), Prof Dorothy Price (Professor of Modern and Contemporary Art and Critical Race Art History, Courtauld Institute), Sara Wajid (co-CEO, Birmingham Museums Trust), Rachael Minott (Inclusion and Change Manager, National Archives), Prof Stephanie Dinkins (Professor of Art, Stony Brook University), Dr Alexandre White (Assistant Prof of Sociology and History of Medicine, John Hopkins University).

Principal Investigator (PI)

Professor susan pui san lok (Professor of Contemporary Art and Director of the UAL Decolonising Arts Institute) is leading the *Transforming Collections* project, directing all research strands, co-leading the public engagement programme, with overall responsibility for the delivery of the project.

Co-Investigators (Co-Is)

Dr Anjalie Dalal-Clayton (Research Fellow, UAL Decolonising Arts Institute) and **Christopher Griffin** (Convenor, Research Programmes and Publications and Editor of Tate Papers, Tate) are co-leading on Strands A&B of the project, focused on surfacing bias across collections and resurfacing artists and artworks. Dalal-Clayton is undertaking an extended and expanded audit of works by black artists in UK collections acquired since 1900. Griffin is developing Tate-focused case studies.

Hilary Knight (Digital Director, Tate; Co-I until June 2022) and **Hannah Barton** (Senior Project Manager, Tate Digital; Co-I from July 2022) led on Tate Collections' Subject Index tagging research, uncovering problematic language in explicit and euphemistic forms, exploring the ways in which this has been addressed over time, and implications for the project research and diverse users' experience (Strands A&B).

Professor Sonia Boyce (Chair in Black Art and Design, UAL; Co-I from July 2022) is contributing to the development of case studies focused on iniva's archive (Strands A&B).

Professor Mick Grierson (Professor and Research Lead, UAL Creative Computing Institute) and **Professor Rebecca Fiebrink** (Professor, UAL Creative Computing Institute) are co-leading on Strands C&D of the project, focused on participatory co-design workshops and ML technology development.

Dr Athanasios Velios (Reader in Digital Documentation at UAL Camberwell, Chelsea and Wimbledon colleges) is contributing to the project technological development with a focus on possibilities, problematics

and limitations of data integration, ontologies and Linked Data, including the potential to record collections' uses of the ML tool (Strands C&D).

Dr Peaks Krafft (Senior Lecturer and MA Internet Equalities Course Leader, UAL Creative Computing Institute) is contributing to ethical considerations of artificial intelligence in terms of conceptualisations of data, and the critical impacts and implications for the technological design, engagement and outputs of the project (Strands C&D).

Mark Miller (Director of Tate Learning) is co-leading Strand E with the PI, focused on public engagement through the project's culminating public programme.

Research Fellows

Ananda Rutherford is developing collaborative case studies on collections information systems and practices, including supporting Dalal-Clayton on the collections audit (Strands A&B), and working closely with **Dr Charlotte Webb** (Senior Lecturer Online courses, UAL Creative Computing Institute) developing and applying data ethics principles in practice across project strands.

Dr Jon Gillick (from July 2022) is focused on the ML development and user experience, from programming and visualisation perspectives respectively (Strands C&D).

Five further Post-Doctoral Research Fellows (from Sept 2022) will develop critical case studies with selected partner collections and organisations, engaging with the research and findings emerging with the evolving ML tool. They are: **Dr Tiffany Boyle**, **Dr Alice Correia**, **Andrew Cummings**, **Dr Tehmina Goskar** and **Dr Ian Sargeant**.

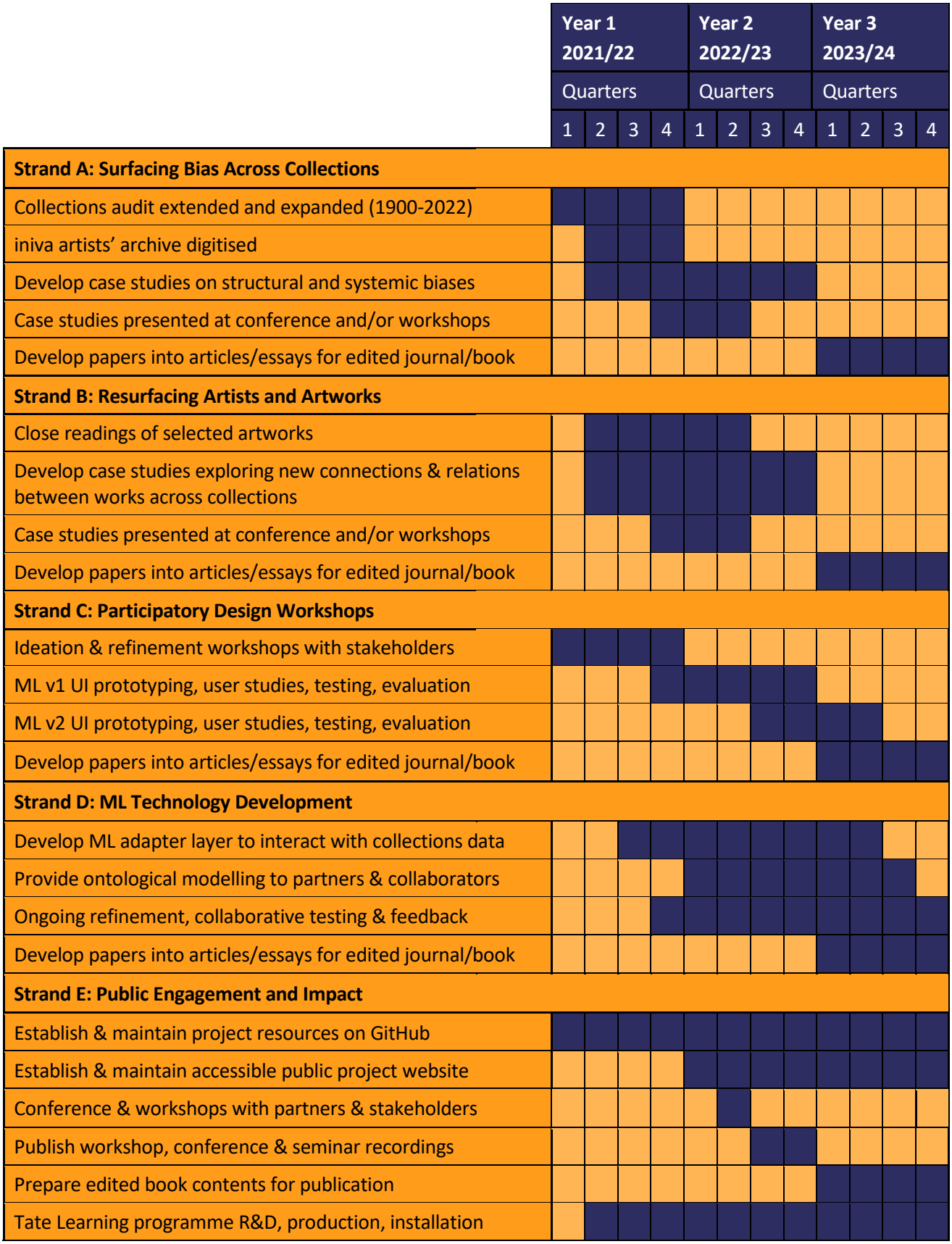
Research Assistant

Kit Bower-Morris is contributing to the early ML prototype development by coordinating and delivering participatory co-design workshops with the team and partners (Strand C).

iniva Archivist

Kaitlene Koranteng (Archivist and Engagement Producer, iniva) is leading on the digitisation of iniva's artists archive.

Overall Programme



Artists practice research residencies in progress												
Artists ML commissions digital build / physical install / live												
Tate installation and public programme live & documented												
Art UK online showcase development, design, production												
Art UK online showcase live												

Events and Consultations

Event / Consultation	**Date**	**Location(s)**	**Attendees**
Audit Consultation	04.02.2022	Online	Tate Information Management team
Audit Consultation	23.02.2022	Online	Cartwright Hall, Bradford
All Day Workshop: Strands C&D	29.03.2022	UAL Decolonising Arts Institute (DeAI)	13 (project team + 1 additional Tate staff)
Tate Collection Research Workshop	30.03.2022	Tate Britain	12 Tate staff (including 2 Tate Co-Is) and 2 TaNC Project team members
Audit Consultation	22.04.2022	Online	Wellcome Collection
All Day Workshop: Strands A&B	27.04.2022	Tate Britain, iniva, UAL Chelsea	16 (project team + 3 additional Tate staff and 2 additional iniva staff)
ML Co-design Conversation	03.05.2022	Online	Wellcome Collection
Audit Consultation	05.05.2022	Online	Sheffield Museums
ML Co-design Conversation	10.05.2022	Online	Arts Council Collection
ML Co-design Conversation	10.05.2022	Online	JISC Archives Hub
ML Co-design Conversation	11.05.2022	Online	British Council Collection
ML Co-design Conversation	11.05.2022	Online	Art UK
Audit Consultation	16.05.2022	Online	Leicester
All Day Workshop: All Partners	18.05.2022	UAL Creative Computing Institute (CCI)	34 attendees (project team, UAL stakeholders, 18 representatives from partner organisations)
Audit Consultation	23.05.2022	Online	British Museum
Audit Consultation	23.05.2022	Online	Herbert Art Gallery, Coventry
Audit Consultation	24.05.2022	Online	National Portrait Gallery
Audit Consultation	24.05.2022	Online	Bristol
ML Co-design Conversation	11.06.2022	Online	Contemporary Art Society
Audit Consultation	20.06.2022	Online	Wolverhampton
ML Co-design Conversation	27.06.2022	Online	Manchester Art Gallery
ML Co-design Conversation	01.08.2022	Online	Birmingham Museums Trust
International Partner Visit	15.09.2022 – 16.09.2022	VAM	Van Abbemuseum (VAM) collections and curatorial staff
All Day Workshop: All Partners	23.11.2022	UAL CCI	TBC
ML Prototype-testing Workshops x 2	03.2023 tbc	Online	c.35 (project team and partners)

Two Day International Conference & Workshops	20.04.2022 – 21.04.2023	VAM	c.150 delegates, public-facing event
Data Ethics Workshops	06.2023 tbc	In-person / online	c.25 (all TaNC Projects)

In addition to the above events and consultations, one-to-one co-design consultations and conversations will continue over the same period. Selected collections site visits will take place over the first half of year two to support the development of case studies.

Research Approach

Context

Transforming Collections seeks to complicate the 'digital cultural record' by embedding decolonial voices and decolonizing impulses to interrogate the structures and narratives of art, nation and heritage. Building on the findings of the AHRC Black Artists and Modernism project (2015-18, led by Prof Boyce with Prof Iok as Co-I and Dr Dalal-Clayton as Research Fellow) and the UKRI MIMIC project (Musically Intelligent Machines Interacting Creatively, 2018-21, led by Prof Grierson), the project also reflects on the recent TaNC Foundation project, *Provisional Semantics* (led by Tate with Dr Dalal-Clayton as Co-I and Rutherford as Research Fellow). The project also builds on UAL and Tate's past and current work on the challenges facing collections. These include 'Tate Encounters' (Dewdney, Dibosa, Walsh, 2007-12) and the HLF-funded 'Archives and Access' project involving digital access, participation and learning with archives; revisions to the 'Art and Artists' database; and an ongoing audit of legacies of slavery in the historic collection (2012-17). Tate's 'Accounts and Accountability' project (led by Griffin) aims to address the history of collection interpretation and the ways in which it has perpetuated racism, imperialism and white supremacy, to ensure that all new written accounts are representative of Tate's commitment to anti-racism, institutional accountability and inclusive language.

Transforming Collections takes up the problem of persistent colonial perspectives and narratives, whereby public institutions (including collections) contribute to and perpetuate an oppressive shared cultural archive[5]. Structural and systemic bias can be discerned across the cultural and technological spaces of the project. Researchers at the intersection of social justice and artificial intelligence (AI) have shown that AI systems can amplify racism, sexism, ableism, and other forms of discrimination; ML can be used to detect and balance bias in text collections[6]; and data augmentation can help to prevent discrimination and disambiguate data structures[7]; while data feminism presumes that oppressive systems of power harm all and hinder the possibility of creating of lasting social impact with data science[8]. Artists, activists and academics (such as Stephanie Dinkins' transmediale practice) continue to build the movement towards equitable and accountable AI (such as the Algorithmic Justice League, founded by Boulamwini, 2016), and ethical treatment of data to combat racial injustice (such as Black Beyond Data, led by Jessica Marie Johnson and Yomaira Figueroa Vasquez).

5 Gloria Wekker, White Innocence: Paradoxes of Colonialism and Race, 2016.
6 Lucas Dixon et al., 'Measuring and Mitigating Unintended Bias in Text Classification', in Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society (AIES '18: AAAI/ACM Conference on AI, Ethics, and Society, New Orleans LA USA: ACM, 2018), 67–73, https://doi.org/10.1145/3278721.3278729.
7 Shubham Sharma et al., 'Data Augmentation for Discrimination Prevention and Bias Disambiguation', in Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES '20: AAAI/ACM Conference on AI, Ethics, and Society, New York NY USA: ACM, 2020), 358–64, https://doi.org/10.1145/3375627.3375865.
8 Catherine D'Ignazio and Lauren F. Klein, Data Feminism, Strong Ideas Series (Cambridge, Massachusetts: The MIT Press, 2020), https://data-feminism.mitpress.mit.edu/.

Approach

Our approach is to evolve an ethical and interdisciplinary braided approach underpinned by decolonial feminist praxes by 1) folding critical art historical and museological research with 2) creative ML development and participatory design and 3) embedding creative practice-based research in collections through artists' activations of interactive ML. We aim to address barriers between collections, audiences and artworks by:

- Enabling cross-search of collections with adaptable ML to enhance the connectivity, discoverability and accessibility of collections for specialist and general audiences alike;
- Enabling researchers and the general public to surface bias and (re)discover those artists and artworks most marginalised within collections, while also ethically engaging and empowering those most marginalised by the institutions that seek to represent them;
- Building on existing systems to enable fluid connections between and across collections;
- Ethically engaging and empowering diverse stakeholders in designing and testing the ML model;
- Generating wide-ranging user and research case studies, and a free, open lightweight ML tool, framework and resources; to be made available to all TaNC projects and related institutions, with potential to roll out beyond the TaNC programme timeframe;
- Establishing an accessible project website to share discoveries with wider audiences;
- Engaging wide-ranging publics through artists' digital commissions and online showcases;
- Opening up new ways of discovering the sometimes uncomfortable stories collections can tell.

Whether a future 'national collection' is realised as a single unified entity, or as a decentralised, distributed and dynamic 'whole', *Transforming Collections* aims to deepen and complexify connections and engender new relationships between objects, collections and the publics they seek to engage and represent. We propose an ethical and inclusive model of cross-collection search and engagement that begins to democratise and decolonise relations of power – not only in terms of how collections' data is defined, held and accessed, but also in relation to the fluid and collaborative production, expansion and contestation of knowledge. The project has the potential to significantly transform the galleries, libraries, archives and museums (GLAM) sector, and thus impact on related disciplines and practices, such as art history and museology. Mobilising ML to surface at scale and at pace the artworks long 'hidden in plain sight', we aim to demonstrate the radical possibilities for transforming collections and reimagining art, heritage and nation.

Impact on Digital Search

Transforming Collections is not attempting to solve the decades-long challenge of developing universal or unifying database and search standards. Rather, we start from the assumption that information systems are never neutral or objective: that there will always be problems of bias, 'misclassification' and inconsistencies in data; and that the so-called 'truth' of labels or categories can always be contested, whether they appear to be factual or speculative, conventional or controversial.

We approach linguistic data as both bridge and barrier to visibility, searchability and value. We seek to dissolve barriers by testing ML's capacity to uncover patterns in data, combined with data science techniques, to reveal and examine latent biases: e.g. the persistence of problematic and offensive language and terminologies that may reflect inherited colonial views and imperialist values, thus perpetuating systemic discrimination and exclusion; or patterns in purchase data that expose networks of power and influence. We aim to build bridges by training ML to search different objects, formats (e.g. texts, images, sounds) and data management systems, thereby enabling unexpected connections between collections, and new interpretative frames, narratives and histories to emerge.

We are working with partners and collaborating organisations who hold or support distinct UK public art collections of varying scales, from large collections deemed nationally significant to smaller yet arguably equally significant and influential municipal or civic collections. Regardless of profile, all hold a strikingly low percentage of works by artists of African and Asian heritage (less than 5%), compared to the representation of the UK population from 'minority ethnic' or 'non-white' backgrounds (c.13%, 2021 Census). Our partners share a commitment to understanding how artists and works have come to be under- or misrepresented, classified, categorised, 'highlighted' or hidden, described, narrativised, and de/valorised within collections and the wider culture. We will:

- Develop adaptable supervised ML software that can connect and augment different datasets, regardless of formats and locations, and thus enable digital cross-search across collections;
- Use selected IML (Interactive Machine Learning) methods and participatory design approaches to work with diverse academic and non-academic stakeholders including TaNC portfolio projects and general audiences, to control the process of determining significance according to individual experience and expertise, e.g. defining labels, designating values and generating bespoke metadata;
- Train interactive ML models to reveal patterns in data within and across disparate datasets; to see what is 'highlighted' or hidden, 'similar' or 'different', under-explored or absent within collections; to surface old/new relationships and create interconnections;
- Discover new and unforeseen links between disparate objects, stories, histories and locations.

Impact on Research Capability

We are bringing decolonial, critical race theory and intersectional feminist approaches[9] to both the ML development and design, and to the analyses of connections surfaced using the evolving ML framework, as braided work strands. A series of case studies focused on works by 'artists of colour' will demonstrate the significantly enhanced and expanded research capability for multi- and cross-disciplinary academics (e.g. engaged in art historical, museological, artistic and curatorial research); as well as the potential benefit to museum and heritage studies, information and archival sciences more widely. The patterns and connections surfaced through ML will enable researchers to identify, contextualise and interrogate the latent colonialism, racism and imperialism of collection terms, concepts and information hierarchies; situate language usage and understandings through time and across locations; and both generate and accelerate new interpretations and radical re-framings of works and collections. In addition to research case studies, a series of digital residencies undertaken by artists of colour will result in commissions that activate the critical and creative potential of ML in transforming how we access, navigate and understand collections. As such, the project seeks to demonstrate the potential to breach physical boundaries and broach new understandings that may fundamentally transform disciplinary methods and understandings of the objects that collections hold.

Crucially, the project will undertake the strategic digitisation of iniva's uncatalogued archive of 5,000 slides and c.200 artists' files. This will establish a rich and vital resource reflecting iniva's championing of 'artists from around the world whose work and ideas' to 'provide new perspectives for Britain's [still] predominantly western-centric view of the visual arts', and iniva's continuing work 'predominantly with British-born and British-based visual artists of African and Asian descent'. This digitisation will have a significant impact on research capability in this area. Together with the artworks and artists uncovered by an extended and expanded audit of UK public art collections (originally undertaken as part of the BAM project), Iniva's digital archive will generate critical comparative datasets to support a wealth of potential case studies to direct the ML development and training.

Impact on Public Engagement

Transforming Collections will embed public engagement at key points in its design and delivery. Across Years 2 and 3, we will hold workshops/seminars, an international conference, and a series of talks and events, to engage wide-ranging academic and institutional stakeholders, as well as specific communities invested or implicated in particular collections. We are building an open-source research-facing GitHub to host resources as they develop (e.g. videos, presentations, case studies, walk-throughs, and tutorials). In Year 2, we will also launch an accessible public-facing website to engage general audiences in project's aims and activities, and the contested and 'potential histories' of collections beginning to emerge through case studies and a series of artists' residencies. In Year 3, *Transforming Collections* will culminate in a major public installation, interventions, and fortnight-long programme of participatory events in the galleries and online, as part of

9 Margaret L. Andersen and Patricia Hill Collins, 'Why Race, Class, and Gender Matter', in Inequality in the 21st Century (Routledge, 2018); Christina Dunbar-Hester, Hacking Diversity: The Politics of Inclusion in Open Technology Cultures, Princeton Studies in Culture and Technology (Princeton, New Jersey: Princeton University Press, 2020).

Tate Learning's Projects and Programmes. This will be significant and celebratory moment for sharing and reflecting on the project's innovations and discoveries. Featuring eight artists' commissions as physical and/or virtual interventions on site and online, the engagement programme will include participatory workshops, talks and discussions on the ethical, critical, experimental, creative ways in which ML can be activated to transform collections. We will also be working with Art UK to showcase selected artists and artworks online and engaging their Art Detective initiative to invite the wider UK public to contribute to existing knowledge around works in collections. We aim to demonstrate what an emerging, distributed yet connected and evolving digital national collection can look like, and in so doing, empower as many people as possible in the collective reimagining of art, heritage and nation.

Project Design and Methodologies

Transforming Collections is designed around five parallel and interweaving work strands.

Strand A: Surfacing Bias Across Collections – Critical Case Studies

Strand B: Resurfacing Artists and Artworks Across Collections – Critical Case Studies

Strand C: Participatory Design (PD) of ML – Ideation and Refinement Workshops

Strand D: Interactive ML Technology Development – Prototyping and Testing

Strand E: Public Engagement and Impact

Collections research case studies (Strands A and B) and ML technology (Strand D) will be developed in tandem, ensuring that critical research and design processes are mutually informed and driven. Embedded participatory design and feedback mechanisms (Strand C) will be established to sustain a collaborative ethical and interdisciplinary approach. Project strands will draw on the collections audit data, and the new digitised data from iniva's archive, along with diverse datasets provided by Tate and our 15 UK partners and organisations. Strands will address the outlined project questions as follows:

In **Years 1 and 2**, work strands will converge through a series of up to 20 workshops/seminars interweaving critical research case studies and ML design, following a participatory action methodology to avoid extractive relationships[10]. Project partners, collaborating organisations and local stakeholders including members of the public will be invited to participate in selected workshops, along with representatives from all the TaNC Discovery Projects.

Year 2 will see eight artists of colour join the project as practice-based researchers in digital residence across partner collections, including Tate. Artists will make critical interventions within and across collections in response to the unexpected connections surfaced by the emerging research case studies, activating the potential of ML as an ethical, critical and creative tool and medium in the creation of new work.

The **mid-point** of the project will be marked by an international conference co-hosted with Van Abbemuseum (VAM), coinciding with their major exhibition, 'Rewinding Internationalism' which will draw on European museum collections (November 2022 to May 2023). The conference will provide an international context for reflecting on the project's progress; situating the notion of a 'national collection' and questions

10 Sasha Costanza-Chock, Design Justice: Community-Led Practices to Build the Worlds We Need, Information Policy (Cambridge, Massachusetts: The MIT Press, 2020).

of access, inclusion and decolonisation in relation to other national and international initiatives (e.g. the NWO/Dutch Research Council-funded project, The Critical Visitor: Intersectional Approaches for Rethinking & Retooling Accessibility and Inclusivity in Heritage Spaces, 2020-25). The event will also foreground artists' practice-based research interventions into collections; and explore the potential of exhibition-making as a critical site of creative research.

By the end of **Year 2**, in addition to an open-source GitHub hosting project resources (e.g. workshop materials, tutorials, walk-throughs), we aim to have established an accessible public-facing project website, featuring video, audio, visual and text documentation. Research case studies in development will shape an edited journal proposal and a book proposal, as well as alternative forms of publication.

Towards the end of **Year 3**, *Transforming Collections* will culminate in high-profile public engagement activity including a major public programme co-curated with Tate Learning, and an online showcase with Art UK. An edited book illuminating practice research residencies and artworks surfaced through the project will be prepared for online and print publication, to extend the project's impact beyond the funded period.

Early Research Results

Strands A&B: Surfacing Bias Across Collections / Resurfacing Artists and Artworks

Strands A and B have focused on three areas of activity: a collections audit, a Tate case study, and digitisation of iniva's artists' archive. Creating an evidential knowledge base to shape a range of emerging case studies, these activities also inform the participatory co-design and ML development in Strands C and D of the project.

The purpose of the **collections audit** is to create a dataset of currently held information on works by Black artists in a UK-wide selection of 33 public collections.[11] The first audit stage updates the lists of modern and contemporary works (initially identified as part of the AHRC Black Artists & Modernism project) by extending the audit period from 1900 to 2022.[12] As of 30 August 2022, data has been collected and processed for 31 collections, in which we have identified 3,338 artworks within the scope of the audit, by 526 individual artists and artist groups. The second stage is underway, expanding on the basic inventory and acquisition data to gather related information, such as biographical data, cataloguing records, descriptive and interpretive texts. The emerging dataset is beginning to show which artists are currently represented (or not) within a sample of UK collections, and the extent and distribution of their representation over place and time. It will further indicate the volume, variety and variability of available (and absent) information, and related interpretation material currently held. The audit dataset forms the basis of a broader data repository for the critical development and analysis of ML algorithms within Strands C and D. This repository will also include data and information from a selection of relevant galleries, libraries, archives and museums (GLAM) sector materials, such as glossaries of terms, collecting, funding and acquisition policies, sector guidance documents, and government reports.

Research case studies will be supported by the critical application of the project's early ML prototypes to the audit dataset, to help reveal patterns over time and place in, for example, acquisition, display and funding decisions, popularity of particular artists, prevalence of particular media, description and interpretation practices, differences in collecting practices within civic and national collections, or between Scotland, Wales, Northern Ireland and England more broadly, and engagement with diversity, equality and inclusion agenda. Case studies may also focus on what, where, when and how problematic concepts have been used to label, classify, categorise, describe and valorise (or devalorise) objects in collections; and how such concepts perpetuate racist, xenophobic, misogynist, heteronormative, white supremacist and ableist views. Critical reflections on issues around the methodology, terminology and problematics of the audit process itself will also constitute a research paper.

The **Tate case study** addresses the above research questions in several ways, engaging Tate Research, Interpretation and Digital teams. Firstly, by surveying and consolidating all existing collection information that is known to perpetuate bias, including critical commentaries from staff and audiences (e.g. in the form

11 For continuity with the AHRC BAM project, the term 'Black artist' is defined as an artist with African, Caribbean, Asian or MENA-region heritage.
12 The BAM audit covered 31 collections. The Transforming Collections audit covers 33 collections so far, including UAL's own Art Collection, and the holdings of 10 project partners and collaborating organisations: Arts Council Collection, British Council Collection, Birmingham Museum & Art Gallery, Government Art Collection, Glasgow Museums, Manchester Art Gallery, MIMA, National Galleries Scotlandd, Tate, Walker Art Gallery, Liverpool.

of offensive or euphemistic language, the absence or erasure of certain histories, and the under-representation or misrepresentation of identities). Secondly, by analysing editorial approaches and interventions within collection texts (e.g. catalogue entries, gallery captions) to address instances of discrimination. Thirdly, by scrutinizing the infrastructures that determine how Tate produces, stores and presents information and makes discoverable the 150,000 collection items on www.tate.org.uk, namely through a process called Subject Index tagging.

An internal review of the ways in which descriptive terms are applied to collection items during the acquisition and cataloguing process, and subsequently data processed and linked, has revealed variances in approaches to Subject Index tagging since its introduction over 20 years ago. Examples of incorrect, problematic and inconsistent tagging also represent risks of reproducing biases and perpetuating harmful experiences for Tate's audiences. By mapping the numerous departments, disciplines, systems and platforms implicated in knowledge capture and production, information management, and the online presentation of Tate's collection, the review makes evident the contingency of cultural knowledge production. The ongoing research will address questions of authorship, ownership and responsibility over the Subject Index tagging process; the extent and approach to any remediation work on tags and texts, examples of which demonstrate the ethical considerations to be embedded in the project's ML development. The Tate case study is key to *Transforming Collections* in addressing how system architectures come to define information practices and collection narratives. As home to 'the national collection of British art from 1500 to the present day', Tate represents a microcosm of the wider GLAM information landscape and makes a significant contribution to the wider cultural digital record.

Digitization of iniva's archive of approximately 200 'artists from around the world' is underway. Founded in 1994 in response to the lack of 'Black and Asian' artists within the mainstream visual arts sector, the digitised archive will facilitate access to a unique research resource. Pragmatically addressing the question of how we might surface suppressed histories, amplify marginalised voices, and re-evaluate artists and artworks ignored or sidelined by dominant narratives, digitisation will also create a dataset to enable comparative critical analyses. Research case studies may focus on the relative presence, absence, representation and interpretation of artists and works within iniva, and across audited collections, as well as the multiplication, duplication or disruption of connections between.

Strands C&D: Participatory Design and Interactive ML Technology Development

The technological approach within Strands C and D of the project has focused on human-centred **participatory co-design as method**.[13] Individual partner consultations and collective participatory workshops have been held to share early research activities, explore design ideas, user needs and experiences, and provide hands-on demos of our first lo-fi ML prototypes. Informed by Strands A and B, the aim is to not only identify patterns in problematic language, but also to recognise allusions to colonial contexts and ideologies. In Year 2, the focus will expand to explore visual modes of search. With initial ML prototypes under ongoing development, the next stage of the project will focus on building large-scale models for testing; creating

13 Peter M. Asaro, 'Transforming Society by Transforming Technology: The Science and Politics of Participatory Design', *Accounting, Management and Information Technologies* 10, no. 4 (October 2000): 257–90, https://doi.org/10.1016/S0959-8022(00)00004-7.

accessible explanations and feedback mechanisms; exploring the possibilities and limits of using linked data to record the use of the ML tool; and co-facilitating working sessions that explore data ethics in practice.

The technological development and design of the project has been guided by the need for **transparency**, **context** and attention to **absences** as fundamental principles**. Transparency** in terms of methodologies – ensuring clarity on how data will be shared, as well as when, where, and how specific ML techniques are deployed, explaining processes and enabling users to explore alternatives. Transparency also refers to the research aim of surfacing the histories and historiographies of objects, which may demonstrate systemic oppression or marginalisation, and how certain voices, bodies and experiences have remained unheard, unseen or erased. Identifying a lack of transparency and accountability (e.g. through dehistoricised, 'neutral', anonymous, authorial voices) could also demonstrate how racist, xenophobic, misogynist, heteronormative white supremacists and ablest views can be inadvertently perpetuated. Comparative analyses of the narrative structure of texts, may reveal shifting hierarchies of value bestowed on different artists and works within collections, as can the mapping of **absences** – that is, the missing, unrecorded or hidden information, or material that remains undigitised. **Context** refers, then, to the need to make explicit the often-Western Eurocentric socio-political, historical and theoretical views reproduced in the interpretation and analysis of data and information framed as knowledge. Context also points the implicit question of the positionality and partiality of project researchers and participants.

Early discussions within the project foregrounded the need for critical, ethical approaches to data. These have led to the creation of two documents, both of which have been shared with our partners: a **Data Sharing Working Guide** (see Annex A) explaining how we intend to use data within the project; and a set of **Data Ethics Working Commitments** (see Annex B). The data ethics commitments comprise a set of principles and prompts to guide the ongoing research, and structure a series of planned internal and partner-facing workshop sessions focused on embedding the principles within practice. We also propose to expand these sessions into an all-day data ethics workshop to engage all TaNC Programme Discovery Projects.

Strand E: Public Engagement and Impact

Strand E has focused on developing the call for practice research residencies to launch in autumn 2022; the international conference programme to take place in spring 2023 in collaboration with Van Abbemuseum; and initial development of the *Transforming Collections x Tate Learning* public programme to conclude the project in autumn 2024. The call will invite artists to undertake virtual practice-based digital research residencies across partner collections. Selected artists will be encouraged to engage with the project's emerging findings, explore connections across collections, and use the ML software as investigative tool and creative medium. The resulting physical and/or virtual commissions will critically and playfully intervene in and disrupt partner collections and displays. Works created will be exhibited physically and/or online as part of a major public programme curated with Tate Learning and showcased online by Art UK.

Project Outputs

Transforming Collections will generate a range of outputs, including:

- Digitisation of iniva's artists archive to establish a new research resource;
- Novel adaptable ML software for dissemination across the programme;
- Critical and interdisciplinary case studies (papers, articles, book chapters);
- An international conference in partnership with Van Abbemuseum;
- A public programme curated with Tate Learning which will include artists physical/digital practice research commissions with selected partner collections;
- Selected artists and artworks showcased online with Art UK;
- An accessible project website for general audiences;
- An edited book showcasing artists' digital residencies and artworks.

Year 1 will lead to the **digitisation of iniva's artists archive** to establish a new research resource. Year 2 will be marked by an **international conference** in partnership with Van Abbemuseum, Years 2 to 3 will see various **critical and interdisciplinary collections case studies** presented / in preparation (including conference papers, journal articles, book chapters), and the launch of an accessible **project website** with video and audio recordings, case studies summaries and reflections. In Year 3, **novel adaptable ML software** will be disseminated to our project partners and collaborators for beta testing, as well as TaNC projects across the programme. The project will conclude with a **public programme curated with Tate Learning** which will include artists physical/digital practice research commissions engaging with partner collections. Selected artists and artworks will also be featured in and **online showcase with Art UK**.

Transforming Collections will culminate in a major public programme featuring eight artists' commissions as virtual and/or physical interventions, in gallery and online. Developed as part of Tate Learning's Projects and Programmes, the engagement programme will include participatory workshops and discussions on the ethical, critical, experimental, creative ways in which ML can be activated to transform collections. The proposed public programme will be structured in three parts: A **Symposium / Seminar** comprising talks, debates and presentations delivered by key stakeholders, to highlight the key ideas, outcomes and outputs of the research project. The event will be aimed at academics, partners, artists within national and international cultural sectors.

The symposium/seminar will be followed by a **Large-Scale Evening Programme** conceived as a multi-disciplinary event across platforms, with media and digital material installations, performance, sonic based work and discursive components. Artists and creatives will contribute to the programme under guiding themes. The event will represent a celebratory launch moment to increase visibility and advocacy for *Transforming Collections'* broader research ideas, propositions and findings, and the interconnected and collaborative nature of the partnership across institutions and networks.

A **Weekend Series** of discursive, performance and participatory events will explore the questions and themes of the *Transforming Collections'* research project, with the aim of engaging a broad and diverse audience. Contributors and audiences will be drawn from a range of research, artistic, academic, public and cultural interest and professions to offer a range of perspectives. The series will focus on connection, sharing critical thinking, exploring potential areas of impact and application, including for future museum practice. Both the large-scale evening programme and weekend series will be aimed at the general public – local, national and international.

Cross Project Collaboration

A Cross Project Workshop proposal addressing data ethics is in development. The proposed workshop will focus on introducing the project's data ethics principles and exploring how these can be applied in practice, addressing, for example, positionality, algorithmic bias and injustice, transparency and extraction. Exploratory conversations have taken place with Dr Sara Perry, Co-Investigator on the Unpath'd Waters Discovery Project, who is leading the development of a similar framework on Shared Values.

Sustainability and Infrastructure

The *Transforming Collections* project will generate a range of digital outputs including through data acquisition and processing, research and user studies, technical experiments, archive research outcomes, and the creation of new ML tools and models. These will be hosted on the UAL Creative Computing Institute (CCI) public GitHub for a minimum of 10 years, with resources to support dissemination to partners for beta testing and potential wider roll out. UAL CCI commits to maintaining all publicly viewable information in line with w3c and uk.gov guidance, maintaining open and freely accessible data where appropriate, and acting as a resource for future projects.

The CIDOC-CRM ontology[14] will be tested for coverage against emerging data structures and if proved adequate, it will be used to map datasets for ingestion at the adapter layer of the system. If the CIDOC-CRM is not considered adequate, then a custom global schema will be created to accommodate variances not represented within the existing common set of fields. Mapping will be done either through 3M Editor (https://github.com/isl/3Meditor) or using custom Python scripts. Prior to transforming datasets, an evaluation of entity reconciliation within a ML context will take place as well as a similar evaluation for vocabulary alignment. The transformed data will be tokenised and ingested into the ML system. Freely accessible open data may include CIDOC-CRM classes and properties as semantically valid tokens to be used in ML experimentation. After augmenting collection data with new tokens featuring ontologically rich statements, the resulting data will be communicated back to the partners and audiences as Linked Data based on the CIDOC-CRM global schema using popular serialisations.

Guidance will be provided on how the CRM can record uses and changes prompted by the ML system, stressing the limitations of both the CRM and ML. Long term data storage will also be through UAL's GitHub enterprise facility. This will mean that research can continue beyond project end, including through further funding and with new institutions and partners, subject to appropriate agreements and available data.

14 Chryssoula Bekiari et al., 'Definition of the CIDOC Conceptual Reference Model', 2022, 238.

Contacts

General enquiries

ContactTransformingCollections@arts.ac.uk

Professor susan pui san lok

Transforming Collections, Principal Investigator

Professor of Contemporary Art and Director, UAL Decolonising Arts Institute

s.lok@arts.ac.uk

Professor Mick Grierson

Transforming Collections, Co-Investigator

Professor of Creative Computing and Research Lead, UAL Creative Computing Institute

m.grierson@arts.ac.uk

Jerneja Rebernak

Transforming Collections, Project & Partnerships Manager, UAL Decolonising Arts Institute

j.rebernak@arts.ac.uk

Fleur Kaminska

Transforming Collections, Project Administrator, UAL Decolonising Arts Institute

f.kaminska@arts.ac.uk

Links

Transforming Collections Project Page:

https://www.arts.ac.uk/ual-decolonising-arts-institute/projects/transforming-collections

Transforming Collections Zotero Library:
https://www.zotero.org/groups/4566230/transforming_collections/item-list

References

Andersen, Margaret L., and Patricia Hill Collins. 'Why Race, Class, and Gender Matter'. In *Inequality in the 21st Century*. Routledge, 2018.

Asaro, Peter M. 'Transforming Society by Transforming Technology: The Science and Politics of Participatory Design'. *Accounting, Management and Information Technologies* 10, no. 4 (October 2000): 257–90. https://doi.org/10.1016/S0959-8022(00)00004-7.

Azoulay, Ariella A. *Potential History: Unlearning Imperialism*. London; Brooklyn, NY: Verso, 2019.

Bekiari, Chryssoula, George Bruseker, Erin Canning, Martin Doerr, Philippe Michon, Christian-Emil Ore, Stephen Stead, and Athanasios Velios. 'Definition of the CIDOC Conceptual Reference Model', 2022, 238. https://www.cidoc-crm.org/sites/default/files/cidoc_crm_version_7.2.1.pdf

Costanza-Chock, Sasha. *Design Justice: Community-Led Practices to Build the Worlds We Need*. Information Policy. Cambridge, Massachusetts: The MIT Press, 2020.

D'Ignazio, Catherine, and Lauren F. Klein. *Data Feminism*. Strong Ideas Series. Cambridge, Massachusetts: The MIT Press, 2020. https://data-feminism.mitpress.mit.edu/.

Dixon, Lucas, John Li, Jeffrey Sorensen, Nithum Thain, and Lucy Vasserman. 'Measuring and Mitigating Unintended Bias in Text Classification'. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, 67–73. New Orleans LA USA: ACM, 2018. https://doi.org/10.1145/3278721.3278729.

Dunbar-Hester, Christina. *Hacking Diversity: The Politics of Inclusion in Open Technology Cultures*. Princeton Studies in Culture and Technology. Princeton, New Jersey: Princeton University Press, 2020.

Sharma, Shubham, Yunfeng Zhang, Jesús M. Ríos Aliaga, Djallel Bouneffouf, Vinod Muthusamy, and Kush R. Varshney. 'Data Augmentation for Discrimination Prevention and Bias Disambiguation'. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 358–64. New York NY USA: ACM, 2020. https://doi.org/10.1145/3375627.3375865.

Smith, Chris, Stuart Hall, Maya Jaggi, and Arts Council of England. Whose Heritage?: The Impact of Cultural Diversity on Britain's Living Heritage : National Conference, Manchester, 1st-3rd November, 1999 : Keynote Addresses. London: Arts Council of England, 1999.

Wekker, Gloria. White Innocence: Paradoxes of Colonialism and Race, 2016.

Further References

Aikens, Nick, susan pui san lok, and Sophie Orlando, eds. *Conceptualism - Intersectional Readings, International Framings. Van Abbemuseum*, 2019. https://vanabbemuseum.nl/en/research/resources/articles/conceptualism-intersectional-readings-international-framings/.

Boyce, Sonia and Dorothy Price eds. Art History; journal of the Association for Art History. 'Rethinking British Art: Black Artists and Modernism', June 2021. https://arthistoryjournal.org.uk/issues/june-2021/.

Dalal-Clayton, Anjalie, and Ilaria Puri Purini. *Doing the Work: Embedding Anti-Racism and Decolonisation in Museum Practice*. London, UK: Contemporary Art Society and UAL Decolonising Arts Institute, 2022. https://www.arts.ac.uk/ual-decolonising-arts-institute/publications-and-podcasts.

Dalal-Clayton, Anjalie, and Ananda Rutherford. 'Against a New Orthodoxy: Decolonised "Objectivity" in the Cataloguing and Description of Artworks – Features – Paul Mellon Centre'. Accessed 5 May 2022. https://photoarchive.paul-mellon-centre.ac.uk/groups/against-a-new-orthodoxy.

lok, susan pui san (2018) *Through the Gate / An(g)archivery.* In: Deviant Practice. Van Abbemuseum, Netherlands, pp. 176-209. ISBN 9 789082 890501

Annexes

Annex A

Transforming Collections: Data Sharing Working Guide

About Transforming Collections

Transforming Collections aims to dissolve barriers between collections by addressing digital search and research capability as entwined priorities, with public engagement conceived as integral to the research development, design and dissemination. The project will develop, model and test critical, creative, ethically grounded and inclusive technological approaches that enable cross-search of collections and increase research capability – by surfacing patterns of bias, revealing hidden or unexpected connections, and opening up new interpretative frames and potential narratives of art, nation and heritage.

Purpose of this working guide

We have created this working guide to facilitate understanding about the use of partner data in the project. **This is not a legal document**. Rather, the working guide is intended for use alongside our partner collaboration agreements. In order to carry out the Transforming Collections research activity and work effectively with our partners, we need to agree the extent of access that we can have to your collections and data. This guide is designed to clarify how we propose to do this, including how we will securely store and safeguard any data shared with us.

Data Sharing Introduction

Access to and use of data is essential for this project, and we would like to assure partners that UAL will handle your data securely, ethically and sensitively throughout. In the context of this document, when we refer to 'sensitive' data we are referring to data that is confidential; or which contains information about people that could impact them unfairly if made public; or personal data defined as 'special category data' within the Data Protection Act 2018.

The quality of this research depends on our ability to access partners' data, so developing a relationship of trust that enables as much data sharing as possible is a priority. At the same time, protecting the privacy of partners and their beneficiaries is of primary importance for us, and key to the successful completion of the research. This working guide outlines our approach to data collection and sharing and sits alongside the agreement we will put in place between you and UAL as guidance only.

What different kinds of data are involved in the project?

There are different types of data involved in this project – i) some of which is already available on partner websites which we have referred to in the below table as 'publicly available data'), and ii) some of which we will need specific permission to access, possibly because it contains intellectual property rights which require consent from the rights' holder; is published under the terms of a restrictive licence; or is not already in the public domain, and we have collectively referred to these in the table below as 'non-publicly available data'. In both cases these might include personal and/or sensitive data. Rather than provide comprehensive definitions regarding what 'data' is, the below table is intended to help illustrate what 'data' we might wish to access to support transparent discussion and sharing.

Publicly available data (available on partner websites) E.g.	**Non-publicly available data** E.g.	**Personal data** (where UK GDPR / UK Data Protection Act 2018 applies) E.g.
• Published catalogue texts, records and image/object captions • Digital catalogue information published online and/or available on enquiry • Online News updates, features and blogs (e.g. new acquisitions/funding awards) • Digitised and publicly available annual reports and collection acquisition reports, collection policy and plan documents	• Collections/object information that is not made publicly available (e.g. acquisition panels and approvals processes) • Statistical information/aggregate data (e.g. collections composition, storage and display information, condition or analytics on online collections, or breakdown by artist gender, etc.) • Digital asset metadata • Information about legacy information systems (e.g. index cards, history files and other archived information relating to the objects/collection)	• Personal information relating to the purchase of works and artist/estate contact details • Data related to provenance • Acquisition grant recipient information • Personal information relating to collections staff (e.g. an archivist's name) • A person identifiable through a photograph or a photographic caption

What data will we collect and store?

- Image, video and sound files
- Text (files, record field data and descriptive text)
- Metadata (data about any digital file, record, image or asset)
- Interaction data from project engagement activities such as workshops (e.g., mouse movements, keystrokes, command histories, file locations and lists)
- Survey data (Survey data will not be collected unless participant information is effectively communicated, and consent has been fully given and recorded prior to participation)

What data would we like to access if possible?

- Provenance data and information
- Acquisition data and information
- Funding data and information
- Information on donations, gifts and bequests
- Exhibition, display and loan data
- Use and engagement data
- Authorities and term lists
- Personal data, as defined in the above table

What data will we generate?

Term	Description	Example
Processed data	Data about collections content that has been fed into image, audio or text analysis algorithms (sometimes referred to as 'features').	An algorithm processes data about images in a collection, such as average colour, range of colours, list of shapes, objects or patterns.
Machine learning (ML) model	A computer program that has been trained to recognize certain types of patterns.	A system is built to recognise certain words, objects or relationships between things in specific kinds of text, images and sounds.
Translation data	Data that can be used to translate or adapt different types of data so they can be more easily accessed.	Information about how different collection repositories are structured is newly formatted to provide search functionality that works across multiple collections.
Document	A digital file	Internal and external reports or research papers are written. Images and audio recordings are made.
New data	New information about collections content produced through machine learning methods	An ML model generates new classifications for artworks or potential interpretations of existing texts
Software	Instructions that tell a computer what to do	An interactive computer system allows collection owners and archivists to select text, images and other data, and use it to generate potentially new information.

There will be layers of human analysis performed on the data. Initial analysis is likely to happen in spreadsheets. Additional analyses could include data visualization, and information contained in academic publications that we co-author with data owners. We will co-publish with data owners as much as possible, and where this isn't possible for any reason, we will request permission to publish before we do so.

Example: A cluster of documents could be brought up by a search and then visualized.

How will we store and control access to your data?

We will offer two options for storing and controlling access to your data depending on how sensitive it is: partners will determine which option they would like to use. Access requires permission from partners. Your partner agreement will allow you to explicitly name those who you wish to access the data and under what authentication process. If those people are not named, they will not be given access.

Platform	Access control
Option 1: GitHub Pro*, hosted outside UAL on private repositories	Research team (UAL and partners) can access repositories after registration and authorisation
Option 2: GitHub Enterprise with Large File Support (LFS), private repositories	Exclusively authorised UAL staff can access repositories using required 2-factor authentication

*GitHub is a development platform that enables developers and companies to host, maintain, and version control their software. The data will be stored under the provision of the data protection agreement between Github and UAL.

All data will be stored in full compliance with UK data protection legislation and as specifically set out in the partner agreement between you and UAL.

How will we handle personal data?

- Only project members involved in conducting studies or surveys will have access to any personal data.
- Only partners responsible for recruitment, organisation and analysis of surveys, discussion or other project activities will have access to information about participants' identities and contact information.

How long will we store data for?

- All data that we agree to maintain as part of the project will be stored on a UAL server for a period of 10 years unless partners specifically ask us to delete their data, in which case we will do this immediately.
- UAL will not charge for access to data as a part of this project.

How will we share data?

- Public data originating from the project will be stored in a publicly available online repository under a Creative Commons (CC) type licence. Specific licence details will depend on the type of content – for example, software will be made available under an open-source software licence (e.g. BSD/MIT). Open-source software is software with source code that anyone can inspect, modify, and enhance.
- No personal data, or commercially sensitive data of any kind will be made publicly available without the explicit consent of the relevant data subjects, and where we determine this publication is essential to the project, informed consent will be sought prior to publication.
- Personal data will not be shared with other partners or with external parties unless a data sharing agreement is in place. Participant data will be fully anonymised before sharing.

Who is responsible for ensuring data is kept secure and private?

Everyone who is allowed access to the data is responsible for ensuring it is kept safe and private. Access will only be available to a specific list of named people. Overall data responsibility rests with the Principal Investigator, susan pui san lok.

How can you request that data be removed?

Please email the project Principal Investigator, susan pui san lok, if you wish to request that any data be removed from our research repositories: s.lok@arts.ac.uk

How will the project manage risk?

You may hold data that you don't feel confident to use or share. For example, you may be concerned about misuse of images or sounds, or unauthorized use or publication of data. We recognise and understand the risks involved and if you have any concerns, will seek to explore potential solutions to support access and use, rather than simply exclude potential data sources. We will provide proper oversight in order to mitigate common risks associated with research of this kind. As outlined above, the way we handle and store your

data will be highly secure and controlled, with a clear line of responsibility for maintaining access and establishing what is the proper use of data.

How will we manage copyright in images and other works (video, sound recording etc)?
This research may unearth images of artworks or other works with out-of-date or no copyright permissions in either or both layers of copyright i.e. the photographic copyright in the image and the copyright in the underlying artwork featured in the image (if applicable). We may also come across works with no identifiable creators, makers or artists. Unless we deem that a copyright exception applies, the work is out of copyright or we have applied for an orphan works licence via the Intellectual Property Office, we will manage risk in accordance with the paragraph above 'How will the project manage risk?'

What ethical considerations will be in place?
Ethical considerations are at the core of this proposal in our research methods, working practices and data use.

To the extent that the project develops its own algorithms for information systems, we will reflexively critique these developments drawing on the expertise of the research team in intersectional feminist AI ethics. We are developing an internal 'Data Ethics Working Principles' document, which will function as a reflective tool for the research team, to be revisited periodically. This will be made accessible to all partners.

More information on UAL-wide processes can be found on the UAL website.

Annex B

Transforming Collections: Data Ethics Working Commitments

This project aims to develop, model and test critical, creative and ethically grounded technological approaches that surface patterns of bias and structural racism, reveal hidden or unexpected connections, enable cross-search of collections, and open up new interpretative frameworks and potential narratives of art, nation and heritage.

What is this document?

To support our work, we have set out a series of data ethics working principles designed to act as a series of reflective prompts that the project team can use at regular intervals throughout the project.

Who is it for?

It is primarily designed for the project team and contributing artists, but is also understood as an iterative working document, to be adapted as needed to establish principles of engagement with project partners and audiences.

What will it help us do?

- Articulate a set of shared values, principles and expectations to guide us through the project
- Go beyond legal contractual requirements to foreground ethical relations and practices
- Assure the people we are working with that we respect and take seriously their rights, beliefs and wishes

- Hold ourselves accountable for the ways in which we use data and ensure these are in line with the project's aims and values
- Clarify what we can/cannot and should/should not do in this project using AI and ML

Why is this important?

This project involves handling data which is sensitive in many regards. It may be entangled in complicated, structurally racist bureaucracies and behaviours, and often violent, colonialist, ableist and patriarchal histories. We want to be mindful that the consequences of working with data unethically are serious: people may be misrepresented or mischaracterized; people may be disempowered rather than empowered; there may be risks of exploitation, or reputational damage; unwarranted conclusions may be drawn by the research; we may damage trust or damage relationships; or we waste people's time.

Our use of data will comply with the law, and we have set out appropriate legal data sharing agreements with all partners. However, we recognize that legal structures themselves can be problematic, or be instrumentalized by institutions who have more power than individuals. In light of this, we think that ethical working principles are necessary in addition to legal compliance.

This document is accompanied by a glossary of terms, including terms relating to data.

What are our working commitments?

1. Reflect on and engage with positionality
2. Work critically and be open to challenge
3. Centre humans not data or research priorities
4. Challenge algorithmic bias and injustice
5. Recognise and refrain from extractive practices
6. Be transparent in intention and behaviour
7. Protect whistleblowers
8. Obtain informed consent

1. Reflect on and engage with positionality

Within the project we all have particular subjective and disciplinary experiences and beliefs that shape our understanding of data. Where we lack expertise or lived experience, we are susceptible to missing or misinterpreting things. This includes how we make decisions about what to prioritise in the project as well as about more granular research questions.

Reflective questions:

- How are my specific experiences, values or cultural position shaping my decisions about how to work with this data? Am I making assumptions or presumptions?
- Could my approach be better informed by exploring other perspectives, knowledge and experience?

2. Work critically and be open to challenge

Throughout the project, it will be necessary to engage in critical discussion of our processes and presentations of data. Our understanding of the outputs of data models or interactions must be based on a critical understanding of the underlying data and its limitations. To avoid an over-reliance on, or

disproportionate trust in algorithms, we should critically question our data, algorithms and human processes. Collection data exists in hierarchies, so critical reflection is required about what kind of information is privileged, by whom, and why.

Reflective questions:

- Have I critically engaged with the data to see if there are forms of racism, colonialism or other forms of oppression embedded within it?
- Where/when/how was the data collected? Through what process, form, etc.?
- What "raw" forms has the data previously existed in?
- What changes have been made to the data over time already?
- Who or what might be absent from the data? How could this absence be accounted for?
- Are the conclusions I am drawing from data critically reflective?

3. Centre humans

Assumptions that data is somehow objective or neutral are prevalent, but data represents and is created by humans who live within hierarchies of power and knowledge. It is important that data is understood as such so that when we work with it, we foreground the ways in which these hierarchies are encoded into data, and how that encoding can be disrupted.

We also assume we will make mistakes and are ready to be humble in the face of calls for accountability.

Reflective questions:

- Who is this work for?
- Who benefits? How?
- Who does it potentially harm?
- Just because I can use this data - should I?
- Am I speaking for others? Am I misrepresenting their stories, experience and views?
- How will I ensure that I am working with data in ways that are consensual, reciprocal, and that understand data as always co-constituted?
- What duty of care will I enact for workshop participants encountering sensitive and possibly harmful data?
- What is the long-term legacy of the project for those involved beyond the core team?

4. Challenge algorithmic bias and injustice

The term 'bias' has specific meanings in the context of computational ethics and machine learning. Algorithmic bias describes how structural societal inequalities are embedded in data and reflected or amplified when this data is used to deploy ML systems. Examples include facial recognition, predictive policing and healthcare algorithms, all of which have been shown to exacerbate racism and cause harm or unequal outcomes for people of colour, and especially black people. As such, we understand that algorithmic bias reflects and perpetuates institutional racism, just as institutional racism produces and reproduces algorithmic bias.

We are aware that AI and ML have limitations. We acknowledge that they are not a 'solution' for structural racism. We expect it to begin to surface connections and patterns that help us to evidence some of the entrenched habits and problematics around the ways in which we understand art.

Reflective questions:

- Does this model identify patterns of racism? Have I considered the intersections of racism with ableism, class discrimination or misogyny?
- What might this model fail to reveal? What are its limitations?
- What can my findings show about broader issues of racism in structures and institutions?

5. Avoid extractive practices

We want our approach to collecting and working with data to be mutually beneficial, rather than extractive. While we cannot ensure that there will be no unintended discriminatory effects on individuals or groups, we will try to keep in mind what the effects might be, on whom, and how we can mitigate the harm and balance the benefits.

For data we collect in research (e.g. interview transcripts or audits), we have standard research ethics procedures and forms that force us to confront and agree on questions about mutual benefit, risks, how data is collected, from whom, what we do with it after collection, how people are credited/remunerated, etc. However, for data within collections (and the people connected to it), these processes do not exist.

Reflective questions:

- Who might be at risk? How?
- To what extent is it appropriate to articulate/design for the benefit of people NOT represented in the data, in addition to people "in" the data?
- How can I minimise risks to people NOT represented in the data, in addition to people "in" the data?
- Am I working with an artist with whom the project team has a pre-existing relationship? What are the potential assumptions / benefits / risks?
- How/when am I engaging with artists? Is there mutual clarity and understanding of expectations?
- Am I using data from people outside the above group? If yes, why?
- What are the implications / benefits / disadvantages for artists potentially 'captured' by the audit?
- What assumptions have we made about benefits to others?
- Am I compromising any intellectual property rights belonging to artists or other participants in my work?

6. Be transparent

Having clear and transparent data policies that respect people's rights, personal values and beliefs is a fundamental requirement addressed through our legal agreements and data sharing working guide. Beyond that, the use of data in ML models can perpetuate biases in the data. This means we need to be transparent - i.e. to show our 'working out' - when it comes to gathering and interpreting data. Being explicit about what we are doing, why and what the issues are disrupts the model of keeping findings ‘behind a curtain’. This is especially important because participants in and audiences for this research may have different degrees of technical understanding.

Reflective questions:

- Have I been clear about how I gathered and interpreted the data?
- Are key documents and decision-making processes publicly accessible?
- Does my work need to be made clearer or require further explanation for different audiences?

7. Protect whistleblowers

While all participating institutions have existing whistleblower protections and policies in place, they may still be difficult to implement in practice and serve the institution more than the individual.

Reflective questions:

- Have I considered potential risks to individuals coming forward as whistleblowers on the project?
- Have I communicated and agreed mitigations and protections with individuals about potential risks?

8. Obtain informed consent

It is important that we critically engage with the concept of consent and ensure that it is freely given, reversible, informed, enthusiastic and specific.
Reflective questions:

- Do I have input and consent from subjects of the data, not just data owners?
- How do the data subjects feel about this work? About the data itself? About the process of the data's collection and storage?
- Am I enabling people to give an informed 'yes' as well as an informed 'no'?

Extending and navigating these commitments with others

Although we have written these commitments with the core project team in mind, we recognise the need to extend these to others involved in the project:

- We aim to engage partners in a dialogue about how these commitments align with their behaviours and practices.
- We aim to engage artists in a dialogue about making work in line with these commitments and navigating issues around data and data ethics.