**The London Eye: A Panopticon of Punishment for the Digital Age**

**Fit to the Call/Contribution to Theme**

Although historians of crime and criminal justice have been early adopters in their exploitation of digital resources, to date they have consulted individual datasets in isolation from one another (often uncritically), and consequently have not been able to exploit the maximum potential of the extensive data available. By developing new techniques for linking together, comparing, and visualising currently separate datasets, this large-scale project will develop new and transferable methodologies for understanding and exploiting complex bodies of genealogical, biometric, and criminal justice data. This process of mapping and using life-course analysis will investigate a central issue of penology and social policy: the relative impacts of different types of punishment on criminal desistance, health outcomes, employment opportunities, and family-life over the longer-term. The newly analysed and conjoined data will make it possible to chart the fortunes of all Londoners convicted at the Old Bailey between the departure of the First Fleet to Australia (1787) through to the death of the last transported Londoner in Australia in the early 1920s. Prisoners kept in London’s burgeoning prison estate can now be identified and followed in newly available digitized prison records, as well as civil datasets (such as the censuses carried out between 1841 and 1911). Convicts sentenced to transportation will be traced through the richly detailed records of the hulks, convict vessels, and prisons in the metropolis and Australia, as well as in birth, marriage and death records. We will trace the criminal London poor through a plethora of digital records recreating a pan-global prism capable of mapping and analyzing their lives at both the collective and individual level. This project reconstructs the story of family formation, desistence and reoffending on a vast and unprecedented scale. In constructing this digital Panopticon of London criminal justice, this interdisciplinary trans-national project will demonstrate the transformative power of innovative digital humanities research to resolve a number of important questions which have long intrigued sociologists, historians, social-geographers, linguistic researchers, economists and criminologists, who have hitherto lacked the tools to complete the job. We will construct tools for unlocking debates across the social sciences and the humanities. In the process we will create new detailed sources of data and the means to process diverse textually-rich longitudinal and cross-sectional nominal record sets. The digital Panopticon can now be realised.

**Research Context**

It would have been impossible to envisage an application on this ambitious scale without previous substantial academic and financial investment from multiple funders. Three existing and successful initiatives lie at the heart of this application: Old Bailey Proceedings Online (www.oldbaileyonline.org); London Lives (www.londonlives.org) and Founders and Survivors (foundersandsurvivors.org). This project brings together the two research teams which created these projects (Shoemaker and Hitchcock; Maxwell-Stewart and Oxley) in order to analyse and link these internationally renowned but currently independent resources. Using life-course methodologies developed by PI Godfrey, we will employ the data to conduct a wide range of enquiries into the lives of Londoners who experienced the criminal justice system – and the lives of those who experienced the Australian convict system – to identify possible relationships between life events, punishments, penal regimes, new social and economic conditions and a large number of different life courses of persistence and/or desistence in crime (Godfrey 2007, 2008, 2010, 2012).

The second vital research context for this project is existing work in the digital humanities. We will exploit some of the data modelling/visualisation techniques developed by the transnational 'Digging into Data Challenge' projects, in particular data-warehousing and semantic analysis methodologies developed for use with the Old Bailey records by the Data Mining with Criminal Intent project (http://crimnalintent.org). These will be built into a simple 'macroscope' for the visualisation of text length, semantics, and structured data, allowing real-time interactive exploration of the datasets. The 'macroscope' will also incorporate faceted browsing and automated clustering. We will also build on the 'automated matching' techniques developed for the London Lives project, where the computer compares a group of names according to specified criteria in one source with a group in another source within a specified time period (<http://www.londonlives.org/formAuto2.jsp>). Since resulting matches still need to be checked manually (labour-intensive work), our aim is to refine this tool and identify the optimum algorithms for successful automated matching.

The third key research context concerns studies of criminal desistance, a critical area of research for modern social policy and criminologists which is now also beginning to harness research in the humanities to bring about more meaningful studies of how people desist from criminal offending. Criminologists have identified several factors which help modern offenders to reform ie. finding employment, forming lasting relationships, having children, and integrating into civil society Godfrey *et al* (2007, 2008, 2010) traced whether these factors were important in previous centuries in England. Similarly, Maxwell-Stewart and others have begun to examine whether the transported convicts who reformed possessed these protective factors (Maxwell-Stewart, 2013). This data has helped develop theories of desistence that can now be tested in this large scale project with much greater sophistication, and in significant quantifiable numbers.

**Research Questions**

By examining the fortunes of Londoners sentenced to either imprisonment or convict transportation, we will test a claim first made by the utilitarian philosopher Jeremy Bentham in 1812 that confinement under surveillance would prove a more effective system of preventing future offending than convict transportation. Using newly linked data, this project will map the lives of the 28,488 Londoners sentenced to transportation to Australia at the Old Bailey between 1785 and 1875 and the 34,565 sentenced to imprisonment in the period 1842–75. This is an ambitious large-scale cradle-to-grave study of 19th century metropolitan and colonial (re)offending that will attract both academic attention, and wider public interest. Effectively two penal systems operated side by side: one that removed offenders to Australia for a minimum of seven years (effectively for life given the cost of a return passage) and a second that used shorter periods of confinement to punish and reform offenders before releasing them back into their London homes and families. Over the course of the 19th century both of these systems were modified. Prison and convict systems both experimented with early forms of probation and licensing; as well silent systems and short, sharp, shocks.

Our intention is to evaluate the impact of these competing systems and modifications by linking large criminal justice datasets, supplemented by demographic and census data, to study the offending histories of convicts from the period before their appearance at the Old Bailey, during their periods of punishment in Australian and British penal institutions, and after their sentences had ended. For a large number of Londoners we will have details of every quarter session or higher court appearance (supplemented by digitized newspaper reports of lower court appearances). For convicts transported to Tasmania and Western Australia or released under the British license system, we have full offending histories on their records. Revealing a new desire to keep convicts under surveillance, both prison and convict record-keeping systems committed intimate details of offenders’ lives to paper. We know the colour of their eyes, names of siblings, height (and often weight) place of birth, former occupation and levels of literacy. Additional electronic datasets allow us to uncover other life details (marriage, employment, residence, birth of children) in order to connect details of private life that correlate to onset into criminality, re-offending and desistence. We will create narratives of the journeys that successive cohorts of ordinary Londoners who were caught up in the criminal justice system (as thousands of people accused of minor offences were) took from cradle to grave. Analysis of records created by the criminal justice system is one of the only means of revealing the life-courses of thousands of working and labouring people who were sentenced to either imprisonment or transportation in this period. As well as charting onset, re-offending, and desistence, it allows us to understand the way that ordinary people interacted with authority and wider structures of governance, as well as showing how structural factors operating at individual level could either initiate or inhibit offending.

By providing the capacity to explore the impact of shifts in penal policy on both the micro and macro scale the project will answer questions of interest to academics, policy makers, and the general public in the UK, Australia and world-wide. In doing so, it will provide concrete demonstration of the huge benefits of using new digital tools for answering conventional historical questions. Moreover, the digital data generated by the criminal justice system also provides substantial bedrock for inter-related important enquiries. For example, using prison and convict records we will explore the longer term inequalities (within and across generations in Britain and Australia) that resulted from involvement in the criminal justice system. We will further develop and interrogate debates in medical/health history which link poor health outcomes with poverty/imprisonment. To what extent, for example, did the impact of a sustained period of imprisonment or enforced migration to a foreign land have on the health of convicts across their lifetimes; and what was the legacy for their children, in terms of not achieving their full height potential, reduced longevity, and increased risk of morbidities.

Explorations of the relationship between crime and communities have tended to use one of two different types of methodology: extensive (quantitative and cross-sectional) or intensive (qualitative and ethnographic). While the first approach offers the possibility of surveying a wide number of variables it provides limited opportunities to explore the nature of individual interactions. Qualitative studies, on the other hand, enable researchers to explore such interactions in much more depth but often encounter difficulties with locating these within wider trends. In order to overcome the limitations of these existing approaches, we need to adopt new digital methodologies which allow a wide range of different datasets to be consulted simultaneously.

The first step is to develop methods for better understanding the data, a pressing issue in the digital humanities. Two interrelated problems at the macro and micro level need to be solved: new methods for visually representing entire datasets need to be developed, and we need better methods for record linkage. Both by comparing the overall distributions of key characteristics of the data in different resources, and by examining how individual convicts and prisoners were represented in each, we will better understand the ideological and practical constraints which shaped each dataset, both the limitations of the original data and those imposed by the processes of digitisation. It is only once these sophisticated techniques for understanding the data (including improved record linkage) have been developed that we can answer our historical research questions. The Georgian and Victorian criminal justice system shaped, and in turn were shaped by, dominant epistemological and ontological frameworks that directed data-gathering. Hegemonic beliefs about criminality and poverty ensured that bureaucratic processes of recording relevant data about convicts `naturally’ privileged some forms of knowledge over others. The way that record sets related to each other, the information excluded/included at various stages of the bureaucratic trail from Old Bailey to Australia was complex, and requires further investigation.

We place ethics at the forefront of our research. Ethics in historical research is a burgeoning area, but very little has been written on the impact of new digital media on ethical research practices. For the first time the democratisation of digital records has brought highly personal data within easy reach of a wide range of lay and professional researchers, and has raised questions about how ethical guidelines can, should, and/or will be applied in the future. The issue of Dark Tourism which is just beginning to be debated has hardly touched upon digital and online data – either in its use as a research tool for individuals, or as a form of presenting visions of the past. We will devote significant research time to try and understand the wider implications of digital Dark Tourism for data-managers, archivists, online researchers, and academics. Our programme of research itself will construct robust ethical practices from the outset, and also reflexively interrogate how ethical perspectives can be maintained and enhanced by this research team.

Our seven research themes described below each have a different focus, but are integrated through mixed-membership research teams and the identification of three theoretical and methodological sets of questions that run through all of the projects.

**How can new digital methodologies enhance understandings of existing electronic datasets and the construction of knowledge?**

\* What can visualisation techniques tell us about the overall shape/distinctive patterns in the data, and what does this reveal about the various processes by which the data were created, and their constraints/limitations?

\* What were the epistemological and ontological constructions which directed the format and formed the limits of information formation in the London and colonial criminal justice systems?

\* How can we improve current record-linkage processes to maximise both the number of individuals linked across different datasets and the amount of information obtained about each individual? What is the minimum amount of contextual information needed in order to conduct successful large-scale record linkage of data pertaining to specific individuals?

**What were the long and short term impacts of incarceration or convict transportation on the lives of offenders, and their families, and offspring?**

\* Which forms of punishment/penal regimes were most effective at achieving deterrence and/or reform; and which appeared to extend criminal careers?

\* How important were factors experienced beyond the prison gates at supporting reform? In other words, who fared best in making a new life for themselves, and why?

\* How long did the impact of convict transportation on health, employment, and relationship-formation last; did it have an intergenerational legacy?

**What are the implications of online digital research on ethics, public history, and `impact’?**

\* How can researchers ethically process such vast amounts of extremely detailed personal data that is so easily and immediately available?

\* What are the implications for public ownership of archives in the digital age; and how can the public and private sector work to ensure that knowledge is democratised through digital delivery?

\* What is the role of digital resources in constructing public histories of crime?

**Theme 1: Epistemologies** (RS lead with TH/DO)

# This project will examine the overall distribution of the data; adopting a 'big data' approach to identify hitherto unrecognised patterns and correlations. Working with structured data and related primary texts, the project will create a range of visualisation environments that allow interactive exploration. For qualitative data, simple measures of text length and data density will be visualised (building on the methodologies developed in the Criminal Intent project), and for structured data faceted browsing and automated clustering methodologies (K-means in the first instance) will be employed to create new ways of exploring the data - of creating what Katy Börner terms a 'macroscope' (<http://dl.acm.org/citation.cfm?id=1897871>). In the first instance, this visualisation environment will be used to identify the ideological and practical constraints which shaped the creation of the data itself. By comparing the information provided about individuals who appear in more than one dataset this project will allow us to critique the image of the individual coded in systematic structures of the record systems we work with. But, as importantly, this project will also allow us to explore how the transformation of these paper records into an electronic form has distorted or altered the meanings we derive from them. This project is an exercise in re-presenting the archive, and in re-assessing the factors that shaped both data collection in the 19th century, and historical writing in the 21st. This is worth doing not only as essential preparation for research on penal outcomes, but also as a way of interrogating how digitisation itself transforms the systems of the knowledge it purports to represent.

**Theme 2: Voices of Authority** (TH lead with RS)

The most personal artefact we possess for most of the men and women and children caught up in the criminal justice system of the eighteenth and nineteenth centuries is their words recorded as testimony, their voices. This project uses two sources of testimony (the *Old Bailey Proceedings* and Hobart Magistrates' Court records) to identify styles and patterns of speech recorded in court, and then to characterise the individuals involved; and finally to relate changing patterns of verbal self-presentation, to defendants' life course experiences. The project will use the individuals identified through nominal record linkage to create a subset of people for whom we have two or more records of spoken testimony. By using Magnus Huber's recently created Old Bailey Corpus (<http://www.uni-giessen.de/oldbaileycorpus/>), in combination with an additional body of Hobart records marked up to the same XML schema, a measure of variations in language will be added to data reflecting life course experience, allowing us to explore how 'character', or at least verbal self-presentation impacted on life changes. Text reflecting testimony will be extracted from the trials, and subjected to measures of vocabulary, word length and structural complexity, using Voyant Tools. Additionally, testimony will be processed using a thesaurus-based, explicit semantic approach that will allow direct measures of sentiment. This builds on the work of Hitchcock in collaboration with Simon DeDeo and Sara Klingenstein at the Santa Fe Institute. Automated clustering methodologies, including but not limited to, Bayesian nested sampling and K-means clustering will be used to explore the resulting data. In the process, the defendants' ability to speak to power in their own voice will form a new measure and data point, through which to understand both the experience of individual defendants, and the factors determining the course of their lives, and relationship to power.

**Theme 3: Penal Outcomes** (BG lead, with HMS, RS)

Recent work has examined the impact of life course events such as migration, employment, marriage and parenthood on offending patterns. The data assembled for this project will for the first time allow these issues to be examined over the long term, while also comparing the efficacy of different penal regimes, with an unprecedented level of detail. We will also explore whether gender was a significant factor in the chances of surviving incarceration/transportation, and remaking their lives post-release. We will calculate reconviction rates/recidivism (one of the most pressing concerns for contemporary/modern criminal justice policy makers); test whether the children of offenders were more likely to offend themselves (intergenerational social transmission); identify structural factors acting at the individual level which supported desistence from crime; explore age-graded data to explore the journey from juvenile to adult offending profiles; sift the data for people who appeared as both complainants and defendants to explore offender/victim overlap; and explore the almost completely un-researched area of the `near-miss offender’. This group who appeared at the Old Bailey were clearly vulnerable to arrest and trial but were found not guilty. We will examine the impact of the trial process on speed of any subsequent conviction they collected. By linking our digital criminal justice data to digitised economic/social data we will explore some new theories on the pace of change in society and its impact on offending/reconviction. The extent of economic change and employment opportunities (higher labour need eases ex-offenders back into employment) has been linked to offenders’ chances of rehabilitation. However, it may be that the key factor is actually the *pace* of economic change (Godfrey 2012). Fast-paced positive changes in a local environment (more jobs, better infrastructure, improvements in physical environment) may accelerate desistence trajectories in individual offenders. It has not been possible to establish a correlation between economic growth and desistence, partly because we have not had sufficient runs of reliable statistical data. However, this project will be able to take advantage of newly available digital resources spanning more than a century to approach this question for the first time.

**Theme 4: Intergenerational Inequalities** (HMS lead with BG/DO)

An unusual feature of this project is that it will capture a large amount of intergenerational data including information on intergenerational offending. This is particularly true for Tasmania where the Police Gazettes reported on offenders tried in all levels of the colonial criminal justice systems. As we will link conviction data for the colonially born to birth, death and marriage registers, we will be able to place offenders within a household context enabling us to use information about parental occupation, location of birth, birth order and family size to explore the impact that familial circumstances had on offending patterns. Using these techniques we will be able to map the frequency of colonial offending and reoffending across the population as a whole allowing the project team to gauge the degree to which the families of transported Londoners were able to escape the surveillant eye of the colonial establishment. The same techniques can be used to measure the rate of intergenerational social mobility charting contrasting outcomes for the descendants of convict and free migrants. In this way we will be able to use the digital data to explore some of the longer-term impacts, not only on individuals, but on family-chains stretching over a considerable period of time – a feature which is further explored in the next theme.

**Theme 5: Biometrics** (DO lead with HMS/BG)

Biomedical sciences intersect with history in the field of anthropometrics. Adult height provides a measure of the conditions that influence physical well-being of a population during childhood while height and weight measurements can be used to explore the recent nutritional experience of a population. This project will use biometric measurements for 42,000 male and 24,000 female convicts and prisoners born in London to examine a) changes in the nutritional status overtime, b) the impact of enforced removal to the Antipodes on juvenile growth and c) the recent nutritional history of prisoners on entry and exit from gaol. A primary goal will be to map nutritional change in London over the course of the late 18th and 19th centuries and relate this to prosecution rates, urban density and the spread of sanitation measures. As the project will accumulate multiple measurements for individuals as they passed between institutions or were arrested multiple times it will also be able to explore changes in height and body mass index over time. Building on a recent finding that transported convicts who accumulated multiple convictions in the colonies were shorter than those who did not reoffend, we will explore the extent to which this pattern was prevalent amongst juvenile as well as adult prisoners. This will enable an assessment of the extent to which measures of childhood deprivation can be used as predicators of future offending. We will also examine the manner in which straightened economic circumstances impacted differently upon men and women enabling us to map the moral economy of the intra-household distribution of food across the metropolis.

**Theme 6: Digital Dark Tourism** (HMS/BG)

We are also conscious of emerging debates on “Dark Tourism” which have gained currency in the last fifteen years as a way of describing the heritage industry’s opening up of former sites of pain and punishment, such as Alcatraz or Port Arthur, for personal pleasure. Where a close involvement between academics and museum experts exists, as they do between convict site managers and academics in Tasmania for example, the benefits are clear to see (<http://www.heritage.tas.gov.au/convict_sites.html>, <http://www.portarthur.org.au/>). In working with the heritage industry, historians play a public role which extends knowledge about their areas of interest to the general public, and also help to construct a socio-historical context which can be used to inform public understandings of crime and offending. Almost all gaol, police and court museums now have websites to advertise their attractions, and some heritage sites have very sophisticated ‘online doorways’. So to some extent we can journey to dark sites of pain and punishment whilst sitting at home in front of the laptop, and there are a few ‘places’ which are nearly or wholly cyber-sites of Dark Tourism. A number of issues arise with the possibility of our website also becoming a site of online Dark Tourism, and we will investigate these issues in conjunction with our selected industry partners with an aim of developing best practice within a rapidly-expanding sector of the British and Australian heritage industry.

**Theme 7**: **Ethics and Digital History** (BG/RS)

The ethical implications of secondary analysis of personal details are beginning to be discussed (Richardson and Godfrey 2003). General guidance on ethical research for researchers can be found online (British Society of Criminology <http://www.britsoccrim.org/codeofethics.htm>; the UK Data Archive <http://www.data-archive.ac.uk/create-manage/consent-ethics>). Defamation legislation which enables people to take legal action in the courts if someone made untrue or harmful statements about them does not relate to deceased persons. Whilst this has been unproblematic in the main (and has allowed historians to largely by-pass ethical panels) the ease of availability of online material (and the scale and intimacy of available detail on thousands of people who have been subject to legal proceedings) has raised particular issues. Dealing with online digital material offers rare challenges, but ones that may become more common in the future. There may be a need for a different kind of ethical practice which considers areas such as how we conceive of participants, whether living or dead. For example, where once researchers or family historians would have spent considerable time researching details of a single person or family, now we can access huge numbers of people captured in online data. Digitised records are mostly reproduced wholesale – the photos of documents reveal all, as do the digital newspaper crime reports. The issue of anonymity, with regard to these kinds of official records, may be a dead letter. We seem to be able to know everything about anyone we choose. Moreover, researchers feel that they know enough about a person and the direction their lives took in order to state with a certain amount of authority the connections between offending and other life events. We should remember, however, that these official documents (useful as they are) can never reveal the whole picture. Humans have complex relationships, impressive imaginations and reasoning abilities, some of which find expression in their actions, making them do some things, and stop doing other things. We will never be in a position to fully understand why human beings acted in certain ways. We must remember that digital sources (in bringing together so much data about one person) can give the illusion that we know more about their lives and motivations, than we actually do. Although the use of online digital material in its infancy, it is likely that this is an issue which will demand resolution in the near future, and we aim to be at the forefront of that debate. Indeed, we aim to be midwives of new processes of ethical online research.

**Research Methods**

The methodologies developed will enable us to link the Old Bailey Online to the Founders and Survivors database and simultaneously integrate records from other cognate, but hitherto less accessible record series including: the convict indents for New South Wales; the hulk registers; admission records for London prisons; and the prison licence registers (digitally available through Find My Past). While linkage to colonial and metropolitan birth, death and marriage data, census and other civil records will be more problematic, the routine collection of information about age, occupation, family and place of residence in criminal justice records will assist this process.

One of the strengths of our research methodology is its ability to link information about the same individual recorded at different points in the life cycle. Such data can be used to chart changing individual circumstances. Thus, height and weight can be used to calculate BMI, gaining an insight into the net nutritional intake in the months immediately preceding institutionalisation. As noted above we will also use digital technologies to compare content enabling us to check the integrity of our source materials by examining variations in the recording of routine data, both between different institutions and over time. Checking for systematic variation in reporting can reveal much about differing bureaucratic processes (whether prisoners were measured with their shoes on or off or whether they were asked to state their last occupation or the trade in which they were most skilled) and selection biases in the convict populations. Changes in legislation, the operation of policing and the courts, and labour market conditions are all likely to impact upon the sex and socio-economic composition of those who were institutionalised. We can monitor such shifts by analysing changes in the reported age, sex, literacy, occupational structure and offences of convicted Londoners. Tabulated census returns will be used to locate prison populations within the wider metropolitan context.

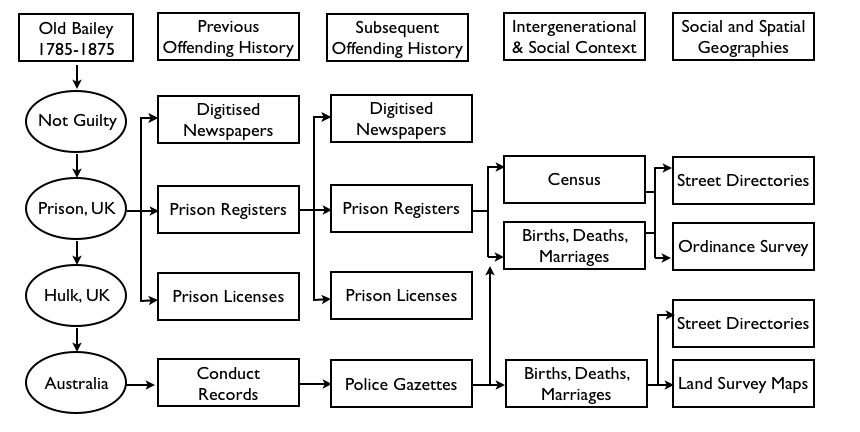
**Summary of Technical Plan**

The technical components of this project involve bringing together approximately 20 datasets that are owned by the project partners, The National Archives, Ancestry, FreeDMB and The British Library. The data represents an estimated 1.3 million individual name instances. The project will develop data visualisation techniques so that the research themes can analyse the datasets both individually and collectively. We will then employ record linkage techniques in order to automatically identify and reconstruct individual lives from the 1.3 million instances. Finally, a website will be developed which will enable academic researchers and the public to interrogate the datasets and generate their own visualisations as well as explore individual lives through the body of evidence. All technical aspects are described fully in the accompanying Technical Plan.

**The research will take place in five stages:**

Stage 1: *Assembling and understanding the data.* From the Old Bailey Online we will select four categories of defendants tried in the period 1785-1875, according to their verdict and punishment: those found not guilty (these will serve as a control group); guilty and sentenced to confinement in prison in the UK; guilty and sentenced to transportation but only served sentence in a hulk in UK; and guilty and sentenced to transportation and sent to Australia. These cohorts will be supplemented with datasets comprising a) all men and women tried in London and landed in Australia 1788-68; b) all convicts landed in Tasmania (1820-53) and Western Australia (1850-68); c) records of subsequent convictions recorded in Australian conduct registers (1820-1875); d) the digitised series of colonial police gazettes (1853-1914); e) digitised Criminal Registers and lists of indictments from quarter sessions including the Hobart Magistrates' Court (some of this material will need to be newly digitised); f) details of previous court appearances recorded in digitised newspapers g) admission registers for Pentonville (1842-1875) and Wandsworth (1859-1875) prisons; h) metropolitan prison licenses (1853-1914); j) metropolitan and colonial birth death and marriage data; and k) British census data. We will add normalised fields and coding structures to all the criminal justice datasets. Visualisation techniques, including both automated clustering and faceted browsing, will then be used to examine the issues raised in research Theme 1. We will create interactive visualisations and identify the overall profiles of these groups and key characteristics, and ascertain any distinctive patterns. To assist with this process and share findings we will conduct a workshop on visualisation techniques involving others with experience or interest in these processes. Stage 1 of the research programme includes the technical work packages WP1 to WP4 (see the Technical Plan).

**Flow chart of research processes:**



Stage 2: *Record Linkage.* Building on the methodologies used in *London Lives*, we will refine procedures and algorithms to determine the most efficient and accurate method of linking names between the different datasets, combining automated and manual processes. These will be recorded using Linked Open Data standards. To facilitate this, we will hold a second workshop, to which those with experience or interest in record linkage will be invited to attend. We will focus our record-linkage efforts between: a) Old Bailey defendants and details of previous court appearances; b) Old Bailey convicts and lists of those transported to Australia; c) transported convicts and colonial birth death and marriage data; d) transportees and records of subsequent convictions; e) transportees and subsequent life-event data (employment, family formation, marriage and child-rearing); f) Old Bailey convicts, prison registers, and metropolitan prison licenses; and g) Old Bailey convicts with birth, marriage and death data. Stage 2 includes the technical work package WP5 (see the Technical Plan).

Stage 3: *Research analysis*. We will address the research questions in Themes 2 through 5 by using life tables, census returns and inter-colonial migration records to model and compare the linked data. This will allow us to examine metropolitan and colonial recidivism rates and explore the impact of different penal regimes on life expectancy, marriage and fertility rates as well as future offending.

Stage 4: *Dissemination*. As discussed below, project findings will be disseminated through workshops and presentations at conferences and seminars; the creation of new online resources; and in print publications. All project publications will be influenced by our investigations into the ethical issues outlined in Themes 6 and 7, which will run concurrently. Stage 4 includes technical WP6.

Stage 5 – *Longer-term research plans*. Using the knowledge and expertise developed over the lifetime of the AHRC part of the research, we intend to carry out further projects on related topics, including further work on biometric data and colonial health records, offenders transported to the American colonies before 1784, prosecution trends in Australia, and the digital mapping of criminal bodies in order to analyse the distribution of scar and tattoo patterns.

**Management and Co-ordination**

The project will be **co-directed** by **Godfrey** and **Shoemaker**, two experienced researchers who have previously managed large-scale Research Council-funded projects, who will work together to ensure that the aims of the project are met on time, and within budget. The **Management Board** consists of all named research leads and also Michael Pidd of the HRI and the Project Manager Sharon Howard. We also have several industry partners which will form an **Industry Partnership Group** (The National Archives, Tasmanian Archives and Heritage Office, ROAR Film, and the London Metropolitan Archives). This will be separate from the **Academic Steering Group** (Prof Peter King, Prof Andrew Prescott, Prof Mark Finnane, Prof Kris Inwood, Dr Rebecca Kippen, Prof, Cormac O’Grada, a PI from one other funded AHRC project, Dr Anita Dockley of Howard League, Liz Hore of National Archives) which will be regularly convened. **Godfrey** will manage the RA in Liverpool. He will be the lead researcher addressing issues of trial procedures, penal regimes, desistance and the impact of imprisonment, and ethics. **Shoemaker** will line-manage the project manager and RA based in Sheffield, oversee the technical work carried out at the Humanities Research Institute (HRI) under Pidd's direction, and be the lead researcher on epistemologies. **Shoemaker** and **Godfrey** will co-supervise the research students. **Hitchcock** and **Shoemaker** will be concerned with digital methodologies and the operation of criminal justice in London**. Maxwell-Stewart** will line-manage the non-AHRC funded Australian-based staff and be the lead researcher concerned with the impact of transportation. Together with **Godfrey**, he will provide expertise in life-grid methodology and whole-life analysis, and co-supervise the Tasmanian based PhD student. **Oxley** will provide vital expertise on 18th and 19th century biometric and health outcomes. **Hitchcock** will lead the specification and evaluation of data visualisations, and implementation of Linked Open Data standards. The research team will meet regularly to coordinate activities, with additional communication taking place electronically and in periodic meetings of the whole project team. The project manager (**Howard**), who managed the Old Bailey and London Lives projects, will run the project on a day-to-day basis, and be the principal point of liaison with the Digital Humanities Developers working at the HRI. The RAs will be responsible for the cleaning and tabulating data, the normalisation and coding of other data files, and checking the results of automated record linkage. Together with the project manager they will be responsible for keeping the blog and website up-to-date. They will fully participate in the dissemination and the publication of research findings. To further train young scholars in the digital humanities, the project team will include 7 research students (two AHRC funded, two funded by Liverpool University, two funded by Sheffield University, and one funded by the University of Tasmania), who will benefit from validated doctoral training (as part of the White Rose, and North West DTCs).

**Timetable**

The funded period of the project lasts for 48 months.

Stage 1: Assembling and understanding the data, including data cleaning, standardisation, coding, documentation and visualisation. Digitisation of additional archives relating to courts, prisons and convicts (as detailed below). (months 1-10)

Stage 2: Record Linkage; this is the most technically challenging part of the project (months 11-24)

Stage 3: Research analysis relating to Themes 1-6 (months 25-32)

Stage 4: Dissemination; including creation of the London Eye website and other electronic resources; writing up books, articles and policy papers; and staging final project conference (months 33-48)

Stage 5: Development of longer-term research plans following this tranche of funding, which will extend the lifetime of the project and allow the full potential of this digital resource to be exploited.

**Outcomes, impact and dissemination**

A strategic dissemination and impact programme will be established. It will stimulate and inform debates at various levels in the following structured and effective way:

At the outset, we will establish a project website, which will include a blog which will encourage two-way discussion and debate about the project's methods and outcomes (social media will also be used for this purpose). The website will document our progress and methodology, introduce issues concerning ethics and epistemology, and include preliminary findings when they become available. In the final year we will create the *London Eye* website, a freely available integrated search engine which will provide access to all project data (and links to externally held data). This major database, containing life course data for 66,000 Londoners who experienced differing penal regimes, will encourage further research by both academics and many of the 12 million family historians in the UK and Australia. To promote better understanding of the data, interactive visualisations will be available and, for those conducting more advanced research, data will be downloadable (copyright issues permitting) via an API. In addition, in partnership with Roar Film and industry partners, we will create online digital learning resources that will disseminate project findings to family historians, schools, and the creative industries.

Project findings will also be disseminated in books and articles intended for academic, tertiary sector and general readerships. All participating universities are committed to open access publishing of the project findings, and authorship will be shared across the research team, including the RAs and PGR students. Two books will be submitted to leading publishers: *The London Eye: Prison versus Transportation* (2017); and *Digital Dark Tourism* (edited collection, 2017). Each research theme will produce substantial articles submitted to high-quality journals (approximately 20 articles) and we anticipate a themed special edition of a history journal.

In order to compound our impact we will organise several workshops, both on methodologies (visualisation techniques and record linkage) and on each of the research themes. We will also organise a final, heavily publicised, project conference, where the empirical results of our research will be presented, demonstrating to a sometimes sceptical humanities audience exemplars of how new methodologies for visualising, linking, and analysing disparate datasets can transform historical research. Other AHRC large grant award holders will be invited to all these events, as will researchers/academics working in this area and archivists/librarians/museum managers and policy-makers/interest groups/activists.

Our impact plan also includes the writing of background papers containing historically informed policy advice aimed directly at policy makers and NGOs, including the Magistrates' Association; Howard League for Penal Reform; Transform Justice; and Prison Reform Trust; and the heritage industry. We will also contribute to popular forums such as the public lecture series of the National Archives and the prestigious Eleanor Rathbone Lecture series at Liverpool University.

**Conclusion**

The digital enterprises of London Lives, the Old Bailey Online and Founders and Survivors have been highly successful projects in their own rights. Each has contributed a considerable amount of data on individuals and longitudinal data on crime rates and sentencing practices. This project will build on this success, enabling new discoveries that will both provide an example to scholars working with other digital resources, and allow historians, criminologists, economists and family historians alike to construct comprehensive life-histories of offenders (pre-, during- and post-conviction) and place these within a meaningful wider context. It maps across time and continents the intergenerational story of crime and desistence, and identifies its impacts on wider family-relationships in London and Australia, in a way which provides data for transformative social research. It will also become a valuable digital repository for an international community of genealogists and non-professional historians providing them with the ability to create their own research projects and place their family stories within a wider and more meaningful social and spatial context.

This project will therefore showcase the way new methods for evaluating and conjoining large datasets can yield research outcomes which, by answering pressing questions, transform a research field. For the first time, it will allow the detailed series of records of crime and justice in eighteenth- and nineteenth-century London and Australia to be exploited by all types of historians to enhance, democratize, and transform our research methods and knowledge-base.

**Jes Text:**

Role Name Organisation Division or Department How many hours a

week will the investigator work on the project?

Principal Investigator

Professor Barry Godfrey

University of Liverpool

Sch of Sociology and

Social Policy

7.5

Co-Investigator

Professor Robert Shoemaker

University of Sheffield History 7.5

Co-Investigator Dr Deborah Oxley University of Oxford History Faculty 3.75

Co-Investigator Professor Tim Hitchcock University of Hertfordshire Social Sciences Arts and

Humanities RI 3.75

**Objectives**

List the main objectives of the proposed research in order of priority [up to 4000 chars]

**Summary**

Describe the proposed research in simple terms in a way that could be publicised to a general audience [up to 4000 chars].

Note that this summary may be published on the AHRC's website in the event that a grant is awarded

There are 4 primary objectives:

1/ To identify and link data about all Londoners tried at the Old Bailey between 1787 and 1875 and subsequently imprisoned or transported to Australia until their deaths (approximately 66,000 men and women).

2/ To create a dataset which links existing online data to newly digitised data and to make that information available from acurated website.

3/ Thereby to create a lasting online digital resource which demonstrates the power of new methodologies in the digital humanities to transform historical research, in this case on the history of crime, sentencing, punishment, convict studies, and desistance for academics, archivists and non-professional researchers.

4/ To investigate the following key research questions: How can new digital methodologies enhance understandings of existing electronic datasets and the construction of knowledge?

\* What can visualisation techniques tell us about the overall shape/distinctive patterns in the data, and what does this reveal about the various processes by which the data were created, and their constraints/limitations?

\* What are the epistemological and ontological constructions which determine the format and limits of information formation in the London and colonial criminal justice systems?

\* How can we improve current record-linkage processes to maximise both the number of individuals linked across different datasets and the amount of information obtained about each individual? What is the minimum amount of contextual information needed in order to conduct successful large-scale record linkage of data pertaining to specific individuals?

What were the long and short term impacts of incarceration or convict transportation on the lives of offenders, and their families and offspring?

\* Which forms of punishment/penal regimes were most effective at achieving deterrence and/or reform; and which appeared to extend criminal careers?

\* How important were factors experienced beyond the prison gates at supporting reform? In other words, who fared best in making a new life for themselves, and why?

\* How long did the impact of convict transportation on health, employment, and relationship-formation last? Did it have an intergenerational legacy?

What are the implications of online digital research on ethics, public history, and `impact'?

\* How can researchers ethically process such vast amounts of extremely detailed personal data that is so easily and immediately available?

\* What are the implications for public ownership of archives in the digital age; and how can the public and private sector work to ensure that knowledge is democratised through digital delivery?

\* What is the role of digital resources in constructing public histories of crime?

Led by an international team of established researchers, this ambitious large-scale project develops new and transferable methodologies for understanding and exploiting complex bodies of genealogical, biometric, and criminal justice data, thereby demonstrating the benefits of digital innovation to sometimes skeptical scholars and public audiences. Through data mapping and life-course analysis this project will investigate a central issue of penology and social policy: the relative

impacts of different types of punishment on criminal desistance, health outcomes, employment opportunities, and family life over the long term. Using sophisticated data-linking methodologies it joins together existing and widely used large data-sets (Old Bailey Online [containing accounts of all trials held at London's Central Criminal Court]; London Lives [a searchable

**Outputs**

**The main outputs of the research**

**Ethical Information**

Provide details of what they are and how they would be addressed [up to 1000 characters]

archive of crime, poverty and social policy]; and Founders and Survivors [records of the 73,000 men women and children who were transported to Tasmania]) with newly digitised data to make it possible to chart the fortunes of all Londoners convicted at the Old Bailey between the departure of the First Fleet to Australia (1787) through to the death of the last transported Londoner in Australia in the early 1920s. Prisoners kept in London's burgeoning prison estate will be identified and followed in newly available digitized prison records, as well as civil datasets (such as the censuses carried out between

1841 and 1911). Convicts sentenced to transportation will be traced through the richly detailed convict records in Australia, as well as in London prison registers and birth, marriage and death records. We will trace the criminal London poor through a plethora of digital records, recreating a pan-global prism capable of mapping and analyzing their lives at both the collective and individual level.

The main output will be a database and curated-website, in addition to publications in leading journals. At the outset we will publish a blog, and a project website (which will explain our methodology and progress, introduce debates on ethics, methodology, and epistemology, and provide preliminary findings, so that we can engage with the widest possible audience). The 'London Eye' website will also provide an integrated search engine for searching the conjoined datasets containing life course data for 66,000 Londoners who experienced differing penal regimes ,which we expect to be regularly consulted by many of the 12 million family historians in the UK and Australia. In addition, in partnership with media and industry partners, we will create online digital learning resources that will disseminate project findings to family historians, schools, and the creative industries. This project reconstructs the story of family formation, desistence and reoffending on a vast and unprecedented scale. In constructing this digital Panopticon of London criminal justice, this interdisciplinary trans-national project will resolve a number of important questions which have long intrigued historians, sociologists, social geographers, linguistic researchers, economists and criminologists, who have hitherto lacked the tools to carry out this research.

Book (co-authored)

Book (edited)

Book (chapter)

Journal article (refereed)

Journal article (non-refereed)

Conference paper

Datasets

Database

Website

Interactive website

Case Study Material

Teaching Material

If Other,

Are there ethical implications arising from the proposed research? Yes

The primary research method is archival research to be conducted at County Records Offices, The National Archives, and, primarily, using online sources. The records we will consult present specific ethical questions to the extent that they contain material that identifies personal details of deceased persons, and research material gained with limited consent which may be re-used. The project will adopt rigorous ethical procedures applying guidance from appropriate professional associations, such as The British Sociological Association and The British Criminology Society. Th PI has previously

published on ethical concerns in historical research and the project management team (alongside the University's research ethics structure) will ensure current standards of ethical governance and protection are applied to the new research. Our commitment to the development of ethical practice in the secondary analysis of online sources can be seen in the Case for Details of where the policy can be accessed

**Academic Beneficiaries**

Describe who will benefit from the research [up to 4000 chars].

Support (i.e. the linked PhD).

Does the institution have a policy on good conduct in research? Yes

Details can be found at: http://www.liv.ac.uk/researchethics/apply,for,research,ethics/#Law

At its broadest, this research will benefit all scholars in the humanities whose research would benefit from the application of new digital methodologies. This includes any scholar using (and seeking to make sense of) large datasets and/or those who seek to compare or link data from different datasets, as well as those interested in interrogating how digitisation itself transforms the systems of the knowledge it purports to represent. More specifically, the project will have specific benefits for scholars in several disciplines:

\*Historians of crime, justice and punishment in England and Australia will come to a better understanding of the impact of eighteenth and nineteenth-century forms of punishment on convicts. They will achieve a better understanding of patterns of recidivism, the impact of different forms of imprisonment (silent system, solitary confinement) and forms of conditional release ('tickets of leave'), and the methods used by colonial authorities to manage transported convict populations. Historians of early Australia will gain a deeper understanding of the life histories of its founding population.

\*Criminologists and sociologists will be able to analyse the role of the numerous contributing factors which can lead to desistance (or not) from crime, both among individuals and, uniquely, across generations. This research will shed light not only on the impact of different forms of punishment, but also of the relative importance of finding employment, forming lasting relationships, having children, and integrating into civil society.

\*Scholars in medical humanities and biomedical sciences will have access to valuable data of the physical characteristics of a large cohort of largely plebeian Londoners over time. They will be able to assess the impact of economic changes and of experiences of punishment (including gruelling transoceanic voyages), both collectively and at the level of individual convicts. They will be able to map nutritional change in London over the course of the late eighteenth and nineteenth centuries, and relate this to economic change, urban density and the spread of sanitation measures, and determine the extent to which measures of childhood deprivation can be used as predictors of future offending. Those interested in gender will be examine how straightened economic circumstances impacted differently upon men and women, and thus map the moral economy of the intra-household distribution of food.

\*Linguistics researchers will be able to analyse the social characteristics of patterns of spoken language in London and Tasmania through analysis of a corpus of xml-tagged trial transcripts, allowing them to examine the evolution of spoken language and sentiment among primarily lower class men and women during a period of significant linguistic change, and trace the early history of spoken English in Australia.

\*Philosophers, social scientists, humanists and those involved in museum and heritage studies who are interested in the ethics of providing instant access to detailed personal information about deceased men and women in the past, both online and in places of 'dark tourism', will be able to participate in debates about the responsibilities of scholars and museums who use such information and make it public.

**Impact Summary**

Impact Summary (please refer to the help for guidance on what to consider when completing this section) [up to 4000 chars]

\*Computer scientists and and information science specialists will be able to contribute to, reflect on, and learn from the innovative methodologies used in the project, including methods for visualising large bodies of data; formulating algorithms for record linkage ; and measuring degrees of linguistic variability and complexity in large bodies of text. An integral part of the project involves engaging non-academic beneficiaries in the use of our data and research findings. The research will contribute to public and policy debates on desistance, penal regimes, and sentencing policy/practice. Our long term aspiration is to bring to a wider audience research on the impact of imprisonment and transportation on the offenders (across the whole of their lives) and their families and communities. The main non-academic beneficiaries of the project are family and local historians, the cultural sector, the heritage industry, schools, and policy makers. In the first month, we will agree an Impact Plan with the steering group and Industry Partners Group. From the start, we will engage with all potential users through the project blog and website and will stimulate debates which connect academic research to

public audiences, as well as encourage and enhance the public use of digital online resources. The economic impact amongst family historians will include driving new subscribers to Ancestry (a model for how humanities research can engage with commercially-driven content providers).We will engage with family historians though contributions to family history magazines, public lectures, and the London Eye. We expect the website to be consulted, like the Old Bailey Online and Founders and Survivors, by many of the 12 million family historians in the UK/Australia, and we hope to harness the

power of the community of family historians for crowd-sourcing. Genealogists researching the site will not only learn about their own family histories but also become aware of the wider history of punishment in their regions, thereby enhancing their quality of life. The research team have on-going relationships/connections with the media and we will seek to reach a larger audience and create more impact for our work by disseminating our findings on radio and television programmes, thereby contributing to the quality of cultural life and the economic health of the cultural sector. Our work with the heritage industry will have similar benefits. As reflected in our research theme on 'Digital Dark Tourism', our investigations into the ethics of the opening up of former sites of pain and punishment for personal pleasure involves issues of considerable interest to that industry. Working with our Industry Partnership Group (comprised of representatives of The National Archives, Tasmanian Archives and Heritage Office, ROAR Film, and the London Metropolitan Archives), we aim to develop

best practice within this rapidly-expanding sector of the British and Australian heritage industry. Beyond these partners, gaol, police, and court museums will benefit from our advice (informal and in briefing papers) and we will present at industry-organized seminars. We will also take the digital archive into UK/Australian schools. Our student-centred approach will allow learners of all levels to navigate their own digital learning experience. Through our contacts with NGOs (including

the Magistrates' Association, Howard League for Penal Reform, Transform Justice, and Prison Reform Trust) and policy makers, and by the dissemination of briefing papers, the project will increase the effectiveness of public services and policy through the provision of a new and innovative perspective on the longer term impacts of imprisonment. Repeat offenders who serve short prison sentences and quickly re-offend upon release cost the UK economy billions of pounds. The

historical perspective provided by this project, and the ability the project has to analyse the impact of different prison and penal regimes on the whole lifetime of offenders can influence policy/public debates. The data will assist the third sector/media to make arguments to policy makers about the efficacy of prisons policy, and to greater engage the public in debates armed with reliable longitudinal digital data.