

DEVELOPING AN R&D AGENDA TO SUSTAIN THE DIGITAL EVIDENCE BASE THROUGH TIME

DEVELOPING AN R&D AGENDA TO SUSTAIN THE DIGITAL EVIDENCE BASE THROUGH TIME

The RecordDNA international multi-disciplinary network is exploring the question 'In the digital era what is the concept of the record and what implications are there for the usability of the evidence base in the future?' This is a grand challenge that requires a range of research and collaborative partnerships to sustain the shifting evidence base through time. This document sets out the case for a formalised research agenda. All of the data presented has been gathered from the RecordDNA workshops and crowdsourcing activities.

MAINTAINING THE EVIDENCE BASE

We all rely on records – to inform decision-making, prove who we are and what we do, research the past, conduct inquiries and hold individuals and organisations to account. As the saying goes in order to develop we 'stand on the shoulders of giants'; we need to know what has gone before. We use evidence today to improve society, our economy and the quality of our governance. We rely on this evidence through time using it for future research and innovation, transparency and accountability. This means we need access to original, authentic and useable records.

However, our digital evidence base is at risk because of the challenges we face in how information is created and captured, how we preserve it as evidence and how we access and use it in the context of constant change. Records are no longer neatly filed together and managed in registries but captured in chains of emails, tweets, other social media platforms etc. They are scattered yet linked. This affects our ability to capture an authentic evidence base as well as to interrogate it and maintain and sustain it. Many copies of a document may exist with unclear authorship, or the definitive original may disappear into a seemingly infinite cyberspace.

What does this mean for the future digital evidence base? Will it contain only evidence that has been recorded and preserved by accident or by absolute (legal) necessity? What kind of evidence base do we want? Is there an 'ideal' and is it necessary to sustain an 'ideal' digital evidence base? Who decides? What value(s) does it have? How does evidence support societies, economies, the planet? Who and how do we fund a sustainable evidence base? These questions, and other ones, highlight the need for a research and development (R&D) agenda to ensure the future digital evidence base is captured, accessible and usable through time. They are not easy to answer. Different stakeholders will have different views and needs; they have different roles to play and expertise to bring to answering them and developing new tools to address the challenges.

THE CASE FOR A RESEARCH AGENDA

In order to maintain the evidence base we cannot wait but must develop and deliver a proactive research agenda which will evolve over time. It is important to deliver solutions that can be applied immediately for current challenges and newly emerging technologies. In addition we need to develop longer term strategies which will need to be tested and trialled over time as well as looking at more fundamental philosophical issues. The definition of R&D below sets out the types of research which need to be undertaken to address this 'grand challenge'.

Research and experimental development (R&D) comprise creative and systematic work undertaken in order to increase knowledge – including knowledge of humankind, culture and society – and to devise new applications of economic, cultural or social value of available knowledge.

Basic research

is experimental or theoretical work undertaken primarily to acquire new knowledge of the underlying foundations of phenomena, observable facts and behaviours, without any particular application or use in view.

Applied research

is original investigation undertaken in order to acquire new knowledge. It is, however, directed primarily towards a specific intended aim or objective.

Experimental development

is systematic work, drawing on knowledge gained from research and practical experience and producing additional knowledge, which is directed to producing new products, experiences or processes or to improving existing products, experiences or processes.

Reference

This definition of R&D is taken from Lomas, E. (2017) Defining R&D for the arts and cultural knowledge domains. London: AHRC and UCL. Lomas built on the OECD. (2015) Frascati Manual 2015: guidelines for collecting and reporting data on research and experimental development. 7th edn. Paris: OECD Publishing. The Frascati Manual definition deals with more rigid forms of science and technology research and this remains embedded within the definition above. However this definition also encompasses wider forms of research which are important in understanding human information needs.

DELIVERING A RESEARCH AGENDA

Research can deliver creative, novel solutions to deal with uncertainty. It can provide specific, generalisable, transferable and/or reproducible solutions. It must connect and build links between different pieces of research. The aim is to develop solutions around the maintenance and use of the evidence base through basic, applied and experimental research, which tackle immediate and future challenges and result in different outcomes and understanding.

Example basic research questions

- What is a record (considering what its forms are today and potential future forms) and what is the ideal evidence base?
- Philosophically how do we understand the existing cultural and societal use of 'evidence' across international and domain boundaries?

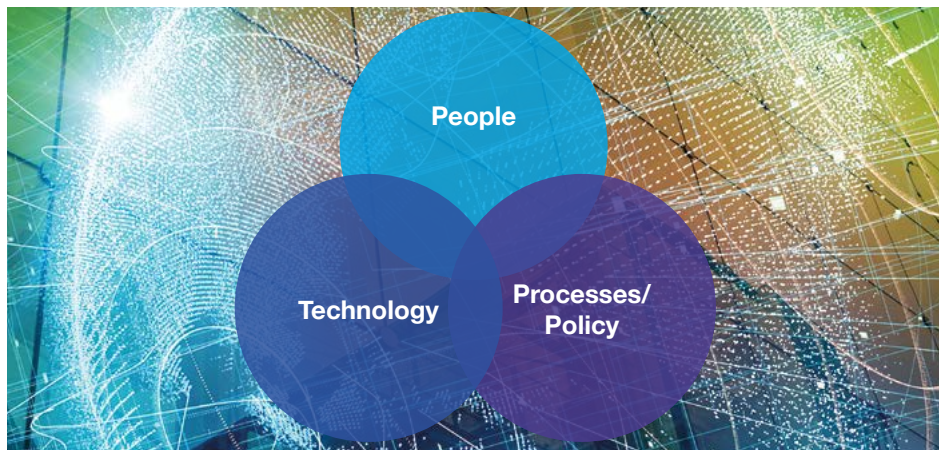
Example applied research questions

- Can we develop business models which enable us to access information through time (in an affordable manner) in particular applied contexts?
- Can we build LIS information literacy frameworks to be applied to archival contexts?

Example experimental development research questions

- How can we better understand the psychology of human interactions with records?
How do we present 'technical proofs' to humans to get humans to trust machine/AI?
Which attributes do people trust and how does this vary with beliefs/culture?
- How do we develop experiments which really approach the problem of managing, preserving and using data and records in quantities beyond human capabilities?

The research agenda needs to resolve issues around people, technology, processes and policy, and their intersection. Underlying all these domains are societal needs, rights and ethics which in some instances are delivered through legal instruments.



RESEARCH THEMES

The research agenda needs to be built across a number of themes with wide ranging associated questions that demonstrate the scale of the challenge, including:

People issues

- Usability and societal needs: How do you ensure that “the record” is created by and useful to all? How do people use evidence? When do people feel the need for evidence? What signs/information creates trust in a group of X users?
- Hidden domains: What ‘evidence’ is out of scope and why, e.g. illegal, dark web, surveillance?
- Research needs: How much research data is available? Can we conduct ethnographic studies of what information the various disciplines want?
- Digital skill sets: Information literacy and digital literacy skills for Archives. Critical data literacy (skills gaps in various user groups).
- Custodians: How much impact does a custodian have on the availability of the digital record?
- Demographics: Do demographics (such as gender) create gaps between custodians, systems and IT developers and influence the digital record?
- Horizon scanning research: to predict future users’ needs and issues.

Technology

- Artificial Intelligence, trust and automation: How do we get humans to trust AI? Does evidence need to be scrutable by humans or is AI better capable of some types of analysis? Does the human factor matter? What should AI do and what should humans do?
- Trust and attacks: How can we trust systems managers? How do we plan the digital evidence base to mitigate for rogue behaviours?
- Encoding and abstraction: How do we encode more of the things we can’t at the moment? What should we encode? How do we take this agenda forward?
- Emulation: Emulation understanding every technical specificity for systems. Do we need to shift from migration to emulation and if so how?
- Algorithms: How does the user know what they are doing? How do we keep users engaged and inspired if it’s all just algorithms doing it for you; that doesn’t encourage engagement?
- Interfaces: Better interfaces/tools for interrogating future digital evidence base. How do we tell the story of a record through our current access tools? How will this happen in the future?
- Multimedia: How much of the record is multimedia? How much multimedia is missed? How are we collecting, storing & preserving content for non-visual and non-textual users?
- Virtual reality: What data are needed to recreate an experience in a VR environment? How do you produce “The Real Thing” – VR experience, the authentic experience?

RESEARCH THEMES CONT.

Policy and processes

- Content: What is content?
- Metadata: How do you manage the metadata that accumulates over time? Can there be a “record” without metadata? Records exist without metadata, but can still be evidence, therefore does metadata necessarily = Record? Think: code?
- Linked data: How do we better link related records? What links already exist?
- Interactions: How do we create records that archive our interactions with them?
- Standards: Can you create standards for the integrity of a record? Can we create standards for the authentication of a record?
- Appraisal and scale: Should we keep everything? Automated systems for appraisal may make processes more transparent? How do we design records that die? How do we best describe all the things you didn't keep? Standards for a record's tombstone.
- Management over time: How do we deal with the time-lapse from record creation to multiple re-use? How do we create systems which preserve information in ways that will be useable over time? How do we deal with perdurance and the continuous curation of inviolate objects?
- Archival processes: How effective are disaster recovery plans for archival resources? How do we make using archival content fun/engaging? Can ethnographic studies of what people enjoy, assist with archival delivery?
- Shifting values: What is the effect of changing purposes/values over time?

Ethics/rights/legal issues

- Ownership: How do we manage the complexities of ownership over time, meanings of provenance and chains of custody, ownership of orphan records in the cloud (who owns them, who can decide what to do with them)?
- Openness: The development of models of openness graded > closed/secure > access permissions > Google crawlers. How do we open more?
- Personal data management and privacy: How do we implement data protection practice into digital records management/curation? How will we manage the right to be forgotten? How might attitudes to privacy change? Is it possible to anonymise?
- Legislative domains: Understanding the existing use of “evidence” internationally, across sectors and culturally. Do access systems provide retrospective juridical contexts/social contexts?
- Ethics: developed balance scenarios to understand conflicts and boundaries such as what is the correct balance between public and private uses for records, e.g. medical records, where they are private records of my medical history, but can be used to assist in treating family members/the community?
- Archival considerations: What entitlement records are needed to use archival resources?

RESEARCH COLLABORATIONS

The research needs to be delivered by global collaborative partnerships that engage a range of stakeholders, from a number of academic disciplines through to practitioners and users, in order to properly draw on the range of expertise required and to meet their varying needs.



A WAY FORWARD

More resource is needed to better connect international multidisciplinary stakeholders in this research agenda and to strategically target connected parts of the research. Timing is critical if we are not to lose our evidence base. Together WE can all tackle the 'grand challenge' of ensuring the future usability of digital records.



RecordDNA

International Research Network

RecordDNA is a Northumbria University and UCL led international research network bringing together academics, practitioners, users and systems developers to build a new cross-disciplinary network of stakeholders with wide ranging expertise to explore the question 'In the digital era what is the concept of the record and what implications are there for the usability of the future the evidence base?' Through this work we hope to aid the maintenance of the digital evidence base through time.

Contacts

Professor Julie McLeod

Northumbria University

Email: julie.mcleod@northumbria.ac.uk

Dr Elizabeth Lomas

University College London

Email: e.lomas@ucl.ac.uk

<https://recorddna.wordpress.com>



Arts & Humanities
Research Council