Est. 1841 | YORK ST JOHN UNIVERSITY

Buil-Gil, David, Zeng, Yongyu, Lu, Yang, Limniou, Maria and Renwick, Robin (2023) Conceptual and Methodological Framework for a Digital Identity and Life-Course Study. Project Report. https://spritehub.org/2023/08/23/digital-identify-and-life-course-study-dialcs/.

Downloaded from: https://ray.yorksj.ac.uk/id/eprint/9729/

The version presented here may differ from the published version or version of record. If you intend to cite from the work you are advised to consult the publisher's version:

Research at York St John (RaY) is an institutional repository. It supports the principles of open access by making the research outputs of the University available in digital form. Copyright of the items stored in RaY reside with the authors and/or other copyright owners. Users may access full text items free of charge, and may download a copy for private study or non-commercial research. For further reuse terms, see licence terms governing individual outputs. Institutional Repositories Policy Statement

RaY

Research at the University of York St John

For more information please contact RaY at ray@yorksj.ac.uk

Sprite+

Security, Privacy, Identity, Trust, Engagement, NetworkPlus

Conceptual and Methodological Framework for a Digital Identity and Life-Course Study

David Buil-Gil[1], Yongyu Zeng[2], Yang Lu[3], Maria Limniou[4] and Robin Renwick[5]

[1]*The University of Manchester*
[2]*Lancaster University*
[3]*York St John University*
[4]*University of Liverpool*
[5]*Trilateral Research*

Table of Contents

Table of Contents 1

Executive Summary 2

1. Introduction 3

2. Review of Existing Studies on Digital Identity in the Life-Course and Conceptual Mapping Exercise 3

2.1 Review of Existing Studies on Digital Identity in the Life-Course 3

2.1.1 Contextual Factors, Perceived Privacy and Online Self-Presentation 5

2.1.2 Individual Factors and Online Self-Presentation 6

2.1.3 Measurement 7

2.2 Conceptual Mapping of Psycho-Social Theories and Digital Identity IT Frameworks 8

3. Proposed Methodology Framework 11

3.1 Research Design 14

3.2 Sampling Approach 15

3.3 Mode of Data Collection 17

3.4 Questionnaire Design and Validation 17

3.5 Data Dissemination, Access and Use 20

3.6 Ethical Considerations 20

4. Way Forward 21

References 23

Executive Summary

The Digital Identity and Life-Course Study (DIALCS) project seeks to set the conceptual, theoretical and methodological foundations for a longitudinal life-course study focused on perceptions of, attitudes to, and behaviours with digital identity technology. This longitudinal study will repeatedly record data from a cohort of participants over a period of time to detect changes in the way they perceive and engage with digital identity technologies. No research has previously examined the adoption and engagement with digital identity technologies over the life-course. Generating such data would be essential not only to better understand citizens' perceptions, attitudes and behaviours towards digital identity, and how these change over time, but also to analyse the impact of emerging and future digital identity technologies in the way people perceive, feel and develop their 'self' identity in digital settings.

The research design of this project is structured in three stages. First, we undertook a rapid evidence assessment of studies on digital identity over the life-course. This was done to identify common themes in the literature, and most importantly, to highlight important gaps in research, which our study will aim to address. Second, we completed a conceptual mapping exercise aimed at linking the most common key terms in psycho-social theories of 'self' and digital identity IT frameworks. This second stage allowed us to identify key constructs that form the core of digital identity, both in psycho-social and technology frameworks. Finally, we ran a series of consultation meetings with domain experts in digital identity and longitudinal research methods. This was done to reach expert consensus on the conceptual, theoretical and methodological foundations for a longitudinal cohort study of digital identity over the life-course. After completing all of these, the following top-level recommendations were reached:

- The study should, where possible, enable descriptive analysis of key terms included in digital identity IT frameworks, government policies, and psycho-social theories of 'self'.
- The study should follow a longitudinal life-course research design.
- The sample size should be large enough to enable population-level estimates and anticipate common attrition issues. Participants will be recruited at the age of 10.
- The sampling approach should follow a stratified random sampling.
- The study should use a combination of computer-assisted telephone interviewing (CATI), computer-assisted personal interviewing (CAPI), and face-to-face interviewing.
- The questionnaire should have two parts: 'general screening form' and 'digital identity form(s)'.
- The study should include measures of use of digital devices, digital access to various platforms, perceptions about digital identity technologies, digital literacy, parental control, experiences with digital technologies, and detailed follow-up questions about the various observed 'digital identities' of respondents.

The impact of the DIALCS will be substantial for scholarly understanding of digital identity, as well as for industry and policy. From an academic perspective, recording longitudinal data on perceptions, attitudes and behaviours with digital identity, both quantitative and qualitative, will enable researchers to address vital questions such as "what drives digital 'self' identity over the life-course?", "how do people construct their 'self' identity in cyberspace?", "what drives people's decisions to engage with certain digital identity technologies but not others?", "what are people's perceptions of security and privacy with respect to digital identity technologies (and how these affect the construction of digital identity)?", and "what indicators represent use and experience of digital identity (e.g., interactions per pseudonym, pseudonym time-lived, etc.)?".

The 'Digital Identity and Life-Course Study (DIALCS): Phase 1' project is funded by the Engineering and Physical Sciences Research Council (EPSRC) through SPRITE+ (grant number EP/S035869/1).

1. Introduction

The objective of the Digital Identity and Life-Course Study (DIALCS) project is to establish the conceptual, theoretical, and methodological underpinnings for a longitudinal cohort study that centres on individuals' perceptions, attitudes, and behaviours regarding digital identity technology. This study will collect data from a group of participants over an extended timeframe, aiming to identify shifts in how individuals perceive and interact with digital identity technologies. Prior to this research, no investigation has explored the adoption and engagement with digital identity technologies throughout individuals' life-course. Generating such data is crucial not only for gaining a deeper insight into citizens' changing perceptions, attitudes, and behaviours concerning digital identity, but also for analysing the influence of emerging and forthcoming digital identity technologies on how individuals perceive, experience, and shape their digital identities.

The research design for this scoping project is structured in three distinct phases. Initially, we conducted a rapid evidence assessment of existing studies on digital identity throughout one's life. This phase allowed us to identify recurring themes in the literature and, most importantly, highlight critical research gaps that our study intends to address. In the second stage, we conducted a conceptual mapping exercise, aiming to establish connections between prevalent terms in psycho-social theories related to the concept of 'self' and digital identity IT frameworks. The primary objective was to pinpoint fundamental constructs that constitute the core of one's digital identity within both psycho-social and technology frameworks. Lastly, we organised a series of consultation meetings with experts in the fields of digital identity and longitudinal research methods. The purpose of these meetings was to obtain expert consensus on the conceptual, theoretical, and methodological foundations for a longitudinal cohort study focusing on digital identity across an individual's lifespan.

This report summarises the main findings of the three phases of our project (i.e., review of existing studies, conceptual mapping exercise, and expert consultation meetings), setting the conceptual, theoretical, and methodological foundations for a longitudinal cohort study on digital identity. The report is structured as follows: First, we summarise the main results of the literature and conceptual reviews. We identify recurring themes and gaps in the literature, and highlight the fundamental constructs that constitute individuals' digital identity. Second, the report summarises the main results of the expert consultation meetings covering the conceptual, theoretical, and methodological foundations for a longitudinal study of digital identity over the life-course. And finally, we present final considerations and ways forward.

2. Review of Existing Studies on Digital Identity in the Life-Course and Conceptual Mapping Exercise

This section summarises the main findings of a rapid evidence assessment exploring the results of previous studies looking at digital identity over the life-course, and the results of a conceptual mapping exercise exploring the core components of individuals' digital identity as highlighted in psycho-social theories of 'self' and IT digital identity frameworks.

2.1 Review of Existing Studies on Digital Identity in the Life-Course

One way to explore 'digital identity' is by answering questions regarding who we are, and who we might be, given current and future interactions with technology. This question delves into what a person is, as well as what a person might become, alongside how we describe ourselves and what we call our 'self'. Individuals may have multiple selves and identities offline (i.e., race,

gender, age, sexual orientation, occupation) (Gaither, 2018), while their online identity may be partly, or even completely, different from their identity in the offline world (Huang et al., 2021) and change over the life-course (Durrant et al., 2015; Friedenberg, 2020).

Higgins (1987) conceptualised individuals' self as actual/real (who they really are in a given situation and at a given point in time), ideal (who they want to be and who they strive to be) and ought to (how they believe others want them to be). Different factors may influence how individuals construct their online selves. These may be based on:

- impressions manipulating their self-presentations (Arkin, 1981; Goffman, 1959 - Self Presentation Theory);
- goals underlying human needs (Katz et al., 1973 - Uses and Gratifications Theory);
- motivations to grow setting goals to better themselves (Deci and Ryan, 1985 - Self Determination Theory);
- their own thoughts, behaviours and feelings to reach goals, based on personal and environmental opportunities and constraints (Carver & Scheier, 1981); and
- the world around them and the different types of computer-mediated communications (Walther, 1992, 1996 - Hyperpersonal Communication Model).

These are only a few of the theories that discuss how individuals construct their 'selves' online. However, in cyberspace, the opportunity for multiplicity exponentially increases, driven by demands for identification and authentication in varied formats, modalities and timeframes. For example, individuals might interact with artificial or extended 'selves' (robot or software), which may bring new opportunities and challenges regarding the sense of belonging, freedom of expression, health treatments, romantic relationships, etc. (Gaggioli, 2017; Mitrushchenkova, 2022; Prescott & Robillard, 2020; Viik, 2020). Individuals might also wish to engage in 'self exploration' online, creating 'short-lived' identities that allow them to explore aspects of their gender, sexuality, or religion with little risk of traceability to their 'real' identity. Individuals might even be incentivised to create identities that persist after their physical being ends. Moreover, such online 'self' identities may vary over time, depending on perceptions about the privacy and security of digital technologies, and technological and regulatory changes.

Digital technologies open up direct concerns related to privacy and security, as aspects such as traceability and linkability become foregrounded (De Hert, 2008; Goodell and Aste, 2019). Such attitudes and behaviours may vary over the life-course or across generations (Orzech et al., 2016). Emerging regulations at the UK (e.g., Digital Safety Bill) and European (e.g., eIDAS2, AMLD6, Digital Services Act, Data Act) level propose stricter requirements for related processes such as age-verification, identity assurance, accountability, and data protection (Belen-Saglam et al., 2023; Schreir et al., 2022). General requirements for identity verification are becoming more widespread, appearing strongly in domains such as public service delivery, insurance, inclusive finance, health, and gaming (Gstrein and Kochenov, 2020; Van Dijck and Jacobs, 2020). The concept of anonymity is being tested, as online services are legally obliged to ensure they know who accesses platforms at any given time (Burke and Moltorisova, 2017). What we do not yet understand is what impact this will have on attitudes and behaviours towards digital identity. Will robust traceability of online interactions cause changes in behaviour (positive or negative)? Will linkability of online interactions promote profiling, censorship, or pave the way for 'black mirror' type outcomes such as persistent social credit scores? Will digital identity pave the way for more ubiquitous and omniscient 'ad-tech' companies to robustly identify individuals, as they traverse the interweb in the pursuit of hugely profitable (and exploitable) 'personal advertisement profiles'?

The longitudinal perspective we are taking to investigate future digital identity is supported by the momentum created by a previous EPSRC-funded explorative study, entitled 'Charting the Digital Lifespan (CDL)'. Findings from CDL provided valuable qualitative insights into how privacy concerns are manifested among new parents and retirees, and the role of social relations

in shaping young adults' building and adding to an online persona (Orzech et al, 2016). Another EPSRC-funded project, 'SID: An Exploration of Super-Identity', developed a concept of digital identity by combining measures across real and cyber domains (Stevenage et al., 2013). While these studies advance our understanding of digital living and the concept of identity, DIALCS is unique in undertaking a longitudinal perspective towards digital identity.

The following subsections describe what previous research has found regarding contextual and individual factors that affect online self-presentation, as well as how other research has measured digital identity in the past. This review of the literature will serve as a foundational element for the DIALCS study.

2.1.1 Contextual Factors, Perceived Privacy and Online Self-Presentation

Previous research has investigated how the design of the digital architecture of social network sites and other online environments may shape dimensions of online self-presentation, including intent, honesty and amount of online interactions. At the outset, the nature of the online sites matters in relation to the type of information published depending on what the socially accepted norms are across contexts (Masur et al., 2023; Stevenage et al., 2013; Van Dijck, 2013). That is, individuals are likely to present aspects of themselves that are 'ideal' for the specific online identity sites. For instance, comparing Facebook and LinkedIn, Van Dijck (2013) argued that the professional nature of LinkedIn would encourage users to present their professional selves in a more controlled manner, whereas Facebook would encourage more personal and emotional content as it is designed to facilitate personal networks. In other words, LinkedIn may discourage self-expression and emotional attachments but encourage self-promotion through recording educational and work experience and achievements, whereas Facebook encourages self-expression to evoke emotions and memories by recording social, emotional and life events in a chronological order. An empirical investigation by Emanuel et al. (2014) into self-statements provided by 148 participants provided evidence on the relation between online contexts and information published. They compared online identities across dating sites, LinkedIn and Facebook, and found that participants reveal less information as the online context becomes more specific. More specifically, job seeking context featured the least disclosure of subjective value-driven information (e.g., 'I am kind') and required knowledge of the person to be verified (e.g., 'I am very sporty'), when compared to the online dating context.

Moreover, it has been argued that the inconsistency between the private and public personas on online sites is not only related to the complexity of human identity and the need for identity management, but also driven by the shifting focus towards engineering and steering social responses through implementing algorithms, especially since 2008, which is reflected in the automated triggering of activities such as 'friending', 'liking', 'connecting' and 'following' (Lee et al., 2010; Van Dijck, 2013). That is, individuals tend to highlight their positive attributes for self-enhancement. Orzech et al. (2017) identified the aim to seek 'likes' among young adults when narrating their photo sharing behaviour as a part of their digital personhood on online social media, and hence the potential mismatch between their 'authentic' and online identity. That is, individuals might stretch the authentic self by, for instance, selectively displaying only flattering photos of themselves and luxury lifestyles (Gibbs et al., 2006).

Interestingly, research has also examined the impact of privacy features of online sites on individual self-presentation patterns. For instance, research by Stevenage et al. (2013) provided evidence that anonymous sites often see more authentic online self-presentation which resemble how people represent themselves in an offline context. This could be attributed to the absence of a well-defined and concrete audience, leading individuals to perceive the sharing of information is safe and there is no necessity to regulate or manage their online self-presentation (Newman et al., 2002; Stevenage et al., 2013). Moreover, research further shows that the online network size and connection type, defined as the number of connections, relations with the followers, affect

self-presentation. For instance, a study by Habib et al. (2019) carried out a survey using a sample of 1,515 Snapchat users and found that many participants shared personal information on Snapchat because they perceived that content was only being shared with a small group of people, which made the content highly personal.

However, the default settings of social network sites may hinder the ability to keep information private. Van Dijck (2013) acknowledged the feature of privacy settings acting as a tool for users to manage their online identities for different audiences. For instance, users may share personal updates with close friends and family while presenting a more professional image to potential colleagues. Despite offering these privacy settings, user profiles are likely to be more public than private because the default is often set to opt out keeping information private. On the other hand, the literature has also argued that online self-presentation is a co-creating process where profile owners and network managers are both in a position to publish content (Van Dijck, 2013). Such loss of control over online identity and concerns about privacy was found to prompt the adoption of strategies to ensure privacy.

2.1.2 Individual Factors and Online Self-Presentation

Research has also shed light on a number of personal demographic characteristics in relation to online self-presentation. Sex has shown to be a predictor of personal information disclosure, where females tend to spend more time on online social networks but place higher privacy restrictions on their profiles than males (Hew, 2011). Race/ethnicity was found to be related to personality and affect online self-presentation (Chen and Marcus, 2012). Literature has also examined the role of personality traits in impacting different dimensions of online self-presentation. Some research demonstrated that extroverts and narcissists are more likely to disclose information about themselves than those who are more introverted (Hew, 2011; Lee et al., 2010). Furthermore, based on a sample of 463 university students, Chen and Marcus (2012) found that dishonest and audience-relevant information in self-presentation (i.e., information that is more relevant to the interests of the audience in an online environment) were predicted by low extraversion and idiocentrism among collectivistic individuals. Similar findings were shown by Hollenbaugh and Ferris (2015), who examined a sample of 301 Facebook users with a mean age of 31.9 years old, and found that honest and accurate self-presentation on Facebook are more common among people who score high on openness as well as exhibitionism (defined as the goal in garnering attention through online sites). Other personal characteristics examined included mental health and previous online experiences. Bessiere et al. (2017) found that depression and low self-esteem are correlated with the higher degree of avatar idealisation where the physical appearance are considered to be more attractive and stand out. Using a sample of 650 adolescents aged 15 to 19 years old, Zimmer-Gembeck et al. (2021) identified that, both online and offline victimisation are associated with online appearance preoccupation, which may increase the discrepancy between online and offline self.

Physical attractiveness has been found to determine whether a humanoid or fantasy avatar is used (Stevenage et al., 2013), and the extent of self-idealisation present in the humanoid avatar. Regarding the use of avatars in virtual environments, while studies have indicated that users tend to adopt avatars that resemble their own gender and appearance, moderate self-enhancement is present when using an avatar in an online world (Jin, 2012; Sibilla and Mancini, 2018). Regarding the use of humanoid avatars versus fantasy avatars to represent the online self, Stevenage et al. (2013)'s study revealed that individuals tend to represent themselves in a manner closely mirroring their real-life appearance when compared to those who use fantasy avatars. This phenomenon may be similarly attributed to the perceived sense of privacy and security afforded by using avatars, or that users see the online world as an extension of the offline world. Moreover, the perceived anonymity of a virtual world has been found to positively correlate with the behaviour of gender swapping when using avatars or the intent to seek interesting experience (Hussain et al., 2008).

Existing literature also highlighted the potential relation between age and user behaviour of digital identity technologies (Brandtzæg et al., 2011; Hollenbaugh and Ferris, 2015; Orzech et al., 2017). Using a relatively small sample of 30 participants and analysing photo sharing behaviour, Orzech et al. (2017) found that most young adults (age 18-23) have Facebook as their primary site for photo sharing with other social media platforms including Instagram, Twitter and Snapchapt being sites for immediate status update; whereas older adults (age 59-70) exhibit online photo sharing behaviour to a lesser extent. Most importantly, it was suggested that self-expression was a distinctive feature among young adults who seek to capture and reflect their identity and share that within their online social networks. Various studies have similarly pointed towards the conclusion that older adults are less likely to use digital identity technologies because they have greater privacy concerns (Quan-Haase and Elueze, 2018; Orzech et al., 2017). Research appears to indicate that the use of privacy settings may further interact with digital literacy. Van den Broeck et al. (2015) carried out an online survey and found that heightened privacy concerns among older adults do not equate to a more frequent use of privacy control features, which may reflect the fact that older adults lack digital literacy to use the privacy feature. For instance, the qualitative study by Brantzæg et al. (2011) revealed that adults over the age of 40 have difficulties in understanding the privacy settings which might make them display completely open public profiles without realising it. On the other hand, younger people, such as college students, are found to selectively use privacy settings to manage their online presentation (Chen and Marcus, 2012). Moreover, Anaraky et al. (2021) recruited 94 participants, including younger adults (19-22) and older adults (65+), to an online web experiment, whose results provided a more nuanced understanding into the interplay between self-presentation, risk beliefs and privacy concerns among older adults. They found that the privacy calculus process among older adults was not only based on data sensitivity, but also the anticipated benefits of disclosure of private information. More specifically, older adults might think more about the goal, risks and long-term outcomes when deciding whether they are to share private information compared to younger adults.

2.1.3 Measurement

We also recorded information about how previous research has measured digital identity and related constructs. The levels of authenticity-self have been previously measured using 5-point POFS Likert-scale using three items (Kurek et al., 2017; Weir and Jose, 2010):

- 'I say what I think even if it is different from the opinions of others.'
- 'I act in ways that express who I really am.'
- 'I can talk openly to others about my feelings.'

Others have measured the intent and honesty of self-disclosure using items included in the Revised Self-Disclosure Scale (Wheeless and Grotz, 1976 cited in Chen and Marcus 2012; Hollenbaugh and Ferris, 2015):

- 'When I wish, my self-disclosures in person are always accurate reflections of who I really am (Intent).'
- 'When I express my personal feelings in person, I am always aware of what I am doing and saying (Intent).'
- 'When I reveal my feelings about myself in person, I consciously intend to do so (Intent).'
- 'In person, I cannot express my feelings when I want to because I do not know myself thoroughly enough (Honesty; reversed).'
- 'I do not always feel completely sincere when I reveal my own feelings, emotions, behaviours, or experiences in person (Honesty; reversed).'
- 'I am honest in my self-disclosures in person (Honesty).'

We also recorded information about related constructs, such as privacy concerns, trust and risk beliefs (Malhotra et al., 2014). Information privacy concerns have been previously

operationalised as the extent to which individuals determine for themselves when, how and how much information about them is communicated to others. Privacy concerns are found to be positively correlated with risk beliefs and negatively with trust beliefs, which influences information disclosure in an online context.

Privacy concern

- 'All things considered, the Internet would cause serious privacy problems.'
- 'Compared to others, I am more sensitive about the way online companies handle my personal information.'
- 'To me, it is the most important thing to keep my privacy intact from online companies.'
- 'I believe other people are too much concerned with online privacy issues.'
- 'Compared with other subjects in my mind, personal privacy is very important.'
- 'I am concerned about threats to my personal privacy today.'

Risk belief

- 'In general, it would be risky to give (the information) to online companies.'
- 'There would be high potential for loss associated with giving (the information) to online firms.'
- 'There would be too much uncertainty associated with giving (the information) to online firms.'
- 'Providing online firms with (the information) would involve many unexpected problems.'
- 'I would feel safe giving (the information) to online companies.'

Trust belief

- 'Online companies would be trustworthy in handling (the information).'
- 'Online companies would tell the truth and fulfill promises related to (the information) provided by me.'
- 'I trust that online companies would keep my best interests in mind when dealing with (the information).'
- 'Online companies are in general predictable and consistent regarding the usage of (the information).'
- 'Online companies are always honest with customers when it comes to using (the information) that I would provide.'

While previous research provides key information about the individual and contextual variables that interact with digital identity, and how to measure key constructs in the digital identity field, the cross-sectional nature of the aforementioned studies mean that they were unable to capture changes in attitudes and behaviour regarding digital identity technologies over the life course. Longitudinal studies can provide valuable insights into the dynamics of the attitudes and behaviour towards digital identity and offer a more nuanced perspective on the relationship between individual characteristics, technological characteristics and development, and users attitudes and behaviours.

2.2 Conceptual Mapping of Psycho-Social Theories and Digital Identity IT Frameworks

Having identified a set of common themes in literature, as well as gaps in research that ought to be addressed in the future, we then undertook a conceptual mapping exercise in order to capture the core elements of individuals' digital identity, as defined in some of the main theories of 'self'

and digital identity IT frameworks. The results of the conceptual mapping exercise are presented in detail in a research article titled *'Identity in the Digital Age: An Analysis of Terminological Disparities in 'Digital Self' Theories and Digital Identity Frameworks'* (Limniou et al., 2023), which is currently under review at a specialised journal. Here we briefly summarise the study and its main findings.

We first selected a set of theories of 'self' that encompass the current landscape of social-psychological frameworks commonly used in the field of digital identity studies. Second, we identified the most pertinent standardised digital identity IT frameworks. Third, we methodically documented the key concepts of each chosen theory and IT framework. And finally, we established connections between the information derived from psycho-social theories and IT frameworks (see analytical strategy in Figure 1). In total, ten psycho-social theories, and five digital identity IT frameworks (i.e., UK DIATF, W3, CENCELEC, EBSI, ISO), were selected, and we recorded information about thirty-five core elements of digital identity considered in at least one framework. A detailed explanation of the analytical decisions taken in each stage of the research process can be found in Limniou et al. (2023).

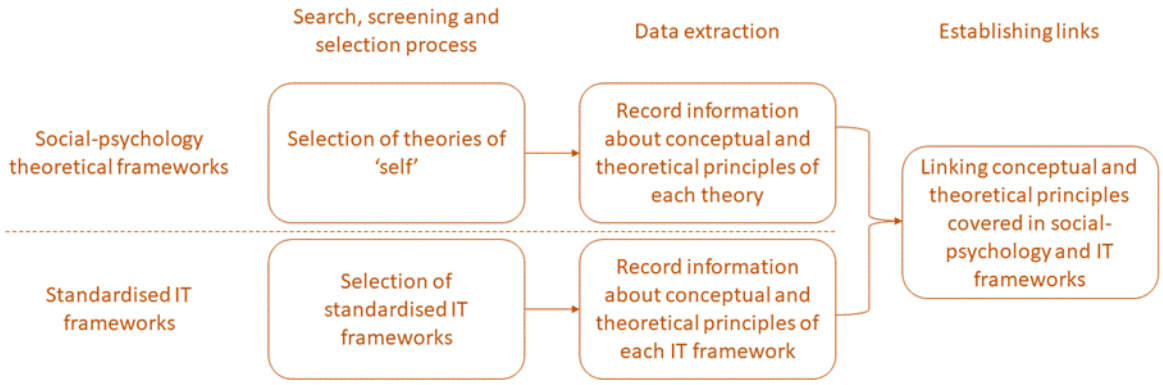

Figure 1. *Analytical strategy of conceptual mapping of psycho-social theories and digital identity IT frameworks*

The following sixteen terms were considered in every single digital identity IT framework included in the study, and at least one of the psycho-social theories, and hence were considered relevant 'constructs' and explored further (i.e., where possible, a life-course study on digital identity should enable at least a descriptive exploration of perceptions, attitudes or behaviours towards each of them):

1. ***Attributes:*** Pieces of information that describe something about a person or an organisation.
2. ***Attribute service providers:*** Individuals or organisations that collect, create, check or share attributes.
3. ***Authenticator:*** Something that users can use to access a service. It could be some information (like a password), a piece of software or a device.
4. ***Cryptographic:*** A way to guarantee the integrity and confidentiality of data transmitted over a public network. This is done by a combination of encryption and signing.
5. ***Digital identity:*** A digital representation of who a user is. Digital identity allows users to prove who they are during interactions and transactions.
6. ***Digital signature:*** A type of electronic distinguishing feature which is used to validate the authenticity and integrity of a message, like an email, credit card transaction or a digital document.
7. ***Encryption:*** Mathematical function that encodes data in such a way that only authorised users can access it.

8. ***Identifier:*** A piece of information that can be used to make a connection between an attribute and a person or organisation.
9. ***Industry standards:*** Relevant established standards from organisations, including but not limited to ISO/IEC, ITU, ETSI, ENISA, ANSI, NIST, and BSI.
10. ***Metadata:*** Data that provides information about other data.
11. ***Orchestration service providers:*** These service providers enable secure sharing of data between participants in the trust framework through the provision of infrastructure.
12. ***Pseudonymisation:*** A security technique that replaces or removes information in a data set that identifies an individual.
13. ***Public Key Infrastructure:*** A way to implement secure electronic transactions over insecure networks. PKI is used to authenticate identities for the purposes of data encryption and signing.
14. ***Relying party:*** Organisations that rely on (or 'consume') products or services from trust framework participants.
15. ***Scheme:*** A group of different organisations that agree to follow a specific set of rules around the use of digital identities and/or attributes.
16. ***Scheme owner:*** An organisation that creates, runs and sets the rules of a scheme.
17. ***Unique identifier:*** Unique data is used to represent someone's identity and associated attributes.
18. ***User:*** People who use digital identity and attribute products and services to prove their identity or eligibility to do something.
19. ***User Agreement:*** Something that confirms users have understood how their digital identities or attributes will be shared.

Other key terms, such as 'data minimisation', 'digital wallet', 'hash function', 'identity service provider', 'personal data store', 'qualified trust service', 'trust framework' and 'zero-knowledge proof' are also commonly used across digital identity IT frameworks and psycho-social theories (details in Limniou et al., 2023), but for the purpose of this report we will focus on the most commonly used constructs. Further research should also consider other terms used across psycho-social and IT frameworks.

Furthermore, we explored which of the considered psycho-social theories of 'self' shared a larger number of relevant terms with each digital identity IT framework (details in Figure 2). Four psycho-social theories presented a higher conceptual convergence with the digital identity IT frameworks and hence their theoretical lens will also be considered in our study (i.e., where possible, a life-course study on digital identity should allow at least a descriptive exploration of the key constructs linked to each theory):

1. ***Self-concept (SC) model*** (Roger, 1959, 1961): Self-concept refers to persons' internal feelings and thoughts and how they view themselves. Self-concept has three components: self-image (how people view themselves), self-esteem (how much value people place on others' worth), and the ideal self (what people wish they were like).
2. ***Self-Discrepancy Theory*** (SDT) (Higgins, 1987): This theory explores how people perceive themselves in both their own and others' eyes, and how others' perceptions of someone are pivotal in how people present themselves to the world. People's behaviour is motivated to reduce the self-discrepancy between the self that they present and the self they ought or wish to be. There are three selves: actual, ideal and ought to.
3. ***Hyperpersonal Communication Model*** (HCM) (Walther, 1992, 1996): Social identity might replace individual identity in an online environment to exaggerate audience impressions by overinterpreting the few cues available that influence the self-presentation strategy. There are four elements that influence this theory: the receiver, the sender, feedback and the asynchronous channels of communication.
4. ***Need-to-belong theory*** (Baumeister, 1999, 2011): Identities are a vital part of the interface between the physical animal body and the cultural social system. A person has

an identity only in relation to other people and other roles. Self-concept refers to the individuals' beliefs about themselves, including the person's attributes and who and what the self is.

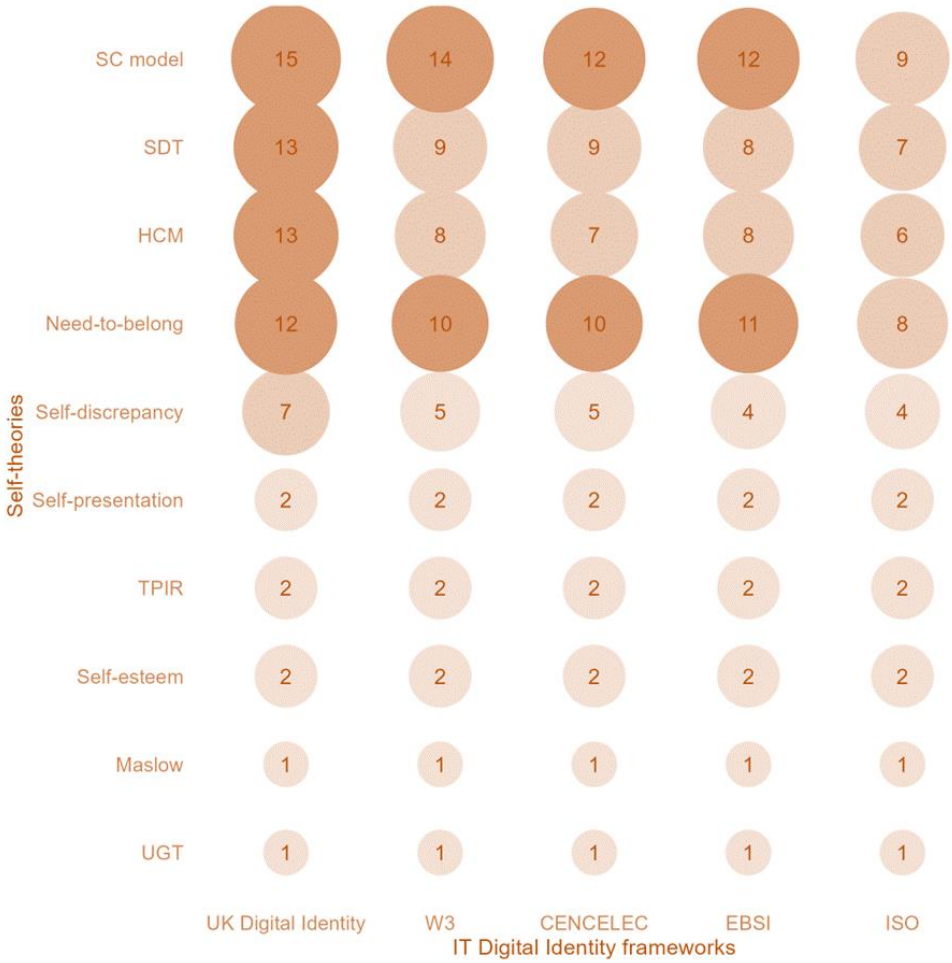

Figure 2. *Conceptual convergence between psycho-social theories and digital identity IT frameworks*

3. Proposed Methodology Framework

We envision DIALCS to repeatedly record data from a cohort of participants over a period of time to detect changes regarding perceptions of, attitudes to, and behaviours with digital identity technologies. No research has yet examined the adoption and engagement with digital identity technologies over the life-course. Generating such data would be essential not only to better understand citizens' perceptions, attitudes and behaviour towards digital identity, and how these change over time, but also to analyse the impact of emerging and future digital identity technologies in the way people perceive, feel and develop their 'self' identity in digital settings.

In order to delineate the conceptual and methodological framework for the DIACS study, we organised a set of consultation meetings with domain experts in digital identity and longitudinal research methods. In total, we organised four consultation meetings, lasting 60 to 90 minutes each. We invited participants based in Academia, industry and the public sector, and asked opinions from diverse groups (based on their gender, ethnicity, expertise, and career stage).

Twenty-seven researchers and professionals were invited to participate, amongst which sixteen accepted the invitation and took part in the meetings. Participants with expertise in Computer Science, Survey Statistics, Human Computer Interaction, Criminology, Cyber Defence, and Data Governance took part in the meetings. Details of participants can be found in Table 1. The consultations meetings took place during July 2023 over video call (Zoom software), and were recorded for coding and analysis by the research team. Three members of the team organised, ran and analysed the content of the consultation meetings.

Table 1. *Details of participants in consultation meetings*

Code	Organisation	Expertise	Sector	Meeting
DAF1	The University of Manchester	*Criminology:* Digital harms, online extremism, identity-based hate	Academia	1
DAM1	University College London	*Computer Science:* Digital currency, digital privacy, decentralised digital identity	Academia	1
DAM2	Alan Turing Institute	*Cyber Defense:* Digital identity, artificial intelligence, computer security	Academia	1
DAF2	The University of Manchester	*Philosophy:* Online harms, hate speech, digital communities	Academia	1
DAM3	University of York	*Computer Science:* Cryptography, digital signatures, verification security	Academia	1
MAM1	Lancaster University	*Statistics:* Longitudinal data, survey statistics, crime	Academia	2
MAM2	The University of Manchester	*Human Computer Interaction:* Multi-item scales, trust measurement, digital identity	Academia	2
MAM3	Utrecht University	*Statistics:* Survey statistics, multi-dimensional constructs, weighting	Academia	2
MAM4	Leeds University	*Criminology:* Survey statistics, measurement error, crime	Academia	2
MBM1	Kantar	*Social Research:* Longitudinal surveys, sampling, fieldwork design	Industry	2
MGM1	Department for Digital, Culture, Media & Sport	*Social Research:* Survey statistics, digital identity	Public sector	2, 3
MGF1	Department for Digital, Culture, Media & Sport	*Social Research:* Survey statistics, digital identity	Public sector	2, 3
DBM1	Linaltec	*Data Governance:* Data privacy, data security, cryptography	Industry	4

For each consultation meeting, we first introduced the project[1] and then probed participants with a set of questions aligned with their expertise and experience. The first and fourth consultation meetings covered substantive questions around digital identity and its measurement, perceptions about digital identity, available data sources, and challenges in studying digital identity. We asked questions such as:

- What are the critical elements of digital identity that should be considered in this research?
- Are there common misconceptions or assumptions about digital identity that we should be aware of?
- How has digital identity evolved over time, and what implications might this have on our study?
- Can you suggest any current research or data sources that might assist us in understanding the dynamics of digital identity?
- What specific demographic or sociocultural factors should be included in the research to ensure a comprehensive understanding of digital identity?
- Are there any specific considerations or pitfalls we should be aware of when studying digital identity in relation to life-course events?
- How can we approach the ethical concerns surrounding the study of digital identities, such as privacy, consent, and data security?
- What are the key challenges and opportunities in studying digital identity?

The second consultation meeting covered practical areas around longitudinal survey methodology, sampling design, data collection, and item validation, including questions such as:

- Given our study's objectives, what research design would you recommend?
- What sampling approach would be best suited for this kind of study?
- At what age should we contact participants for the first time?
- How can we ensure the reliability and validity of our variables of interest?
- What data collection methods would you recommend for a study of this nature?
- How can we account for potential biases or confounding factors in our study?
- What are some strategies for handling missing or incomplete data?
- What kind of data analysis methods should we consider?
- How can we ensure that the research process following ethical principles, especially with regards to data collection, storage, and sharing?

[1] Text read to participants at the beginning of each consultation meeting:
"The Phase 1 of the study, entitled DIALCS, involves researching how people present themselves online. The team wants to understand how people present who they are, who they intend to be, and who others think they are within an online environment. This project will map different theories of online presentation drawn from the domain of social psychology, to understand what they have in common with DI technological solutions. The project team understands that currently developing DI standards either help or hinder self-expression, with the potential impact on aspects of their 'self', especially with respect to trust, privacy, and anonymity over a person's lifetime. The analysis in phase 1 will provide the theoretical basis for the development of the methodology for a longer-term study (phase 2). This long-term study will seek to understand how people integrate DI technologies into their daily online experience, and more specifically how DI technologies support or suppress types of online behaviour. The DIALCS project will include stakeholder involvement to validate findings from the study, as well as to provide critical input into the development of the method for the long-term study in Phase 2."

- Can you suggest any tools or software that can aid in data collection, management, and analysis for this kind of research?
- How can we design the study to ensure it is reproducible and transparent for future research?

The third consultation meeting more directly addressed the feasibility of undertaking the DIALCS study in the future (i.e., DIALCS Phase 2).

The following subsections summarise the main results of the consultation meetings, including its proposed research design, sampling approach, mode of data collection, questionnaire design and item validation, data dissemination, and ethical considerations.

3.1 Research Design

A longitudinal research design will be followed to record data from a sample of participants since their late childhood until their adulthood. A longitudinal research design is a type of research methodology commonly used across various research fields, such as psychology, sociology, pedagogy, and epidemiology, to study changes in individuals or groups over a period of time. Unlike cross-sectional research, which gathers data from a single point in time, longitudinal studies involve repeated data collection from the same participants over an extended period. This approach allows tracking and analysing changes, developments, and trends that occur over time, providing insights into the dynamics of recorded variables. Longitudinal studies involve collecting data from the same participants at multiple time points. This could span weeks, months, years or decades, depending on research objectives. Longitudinal research is well-suited for exploring cause-and-effect relationships because it allows examining how changes in one variable influence changes in another variable over time.

In DIALCS, data will be recorded through questionnaires once a year, in waves. The research design will involve repeated observations of the same variables from the same group of respondents, although the study may also incorporate new questions in later waves if it is deemed appropriate (e.g., questions about new technology developments, online systems, digital platforms). Longitudinal research designs are common across the health and behavioural sciences, but also in the social sciences. In the UK, some examples of longitudinal studies that will be taken as reference points are the Millennium Cohort Study run by the UCL Centre for Longitudinal Studies (https://cls.ucl.ac.uk/cls-studies/millennium-cohort-study/), the Understanding Society study (formerly known UK Household Longitudinal Study; https://www.understandingsociety.ac.uk/) or the more recent COSMO study (https://cosmostudy.uk/about/about-the-study). All these studies have in common that they record data from national cohorts of participants from an early age and focus on a variety of social and behavioural outcomes such as education, deprivation, mental health, inequality, wellbeing, and social mobility.

Longitudinal surveys have also been used to record data about UK citizens' attitudes and behaviours in relation to culture, media and sport, such as the 1970 British Cohort Study (https://cls.ucl.ac.uk/cls-studies/1970-british-cohort-study/), which was run by the UCL Centre for Longitudinal Studies. This longitudinal study serves as one of the most important sources of evidence in understanding what factors can influence British people's 'chances' in life. In the

context of digital studies, there are a number of pedagogical and educational studies that have employed longitudinal research designs in investigating the utility of digital devices in a classroom setting (e.g., Li et al., 2022). Longitudinal designs enabled these studies to better establish and measure the effectiveness of digital technology in learning. Longitudinal designs have also been used to explore identity formation (e.g., Meeus, 2011), though not yet in the digital realm. DIALCS will build on all these different research endeavours with the ambition of repeatedly recording data from a cohort of participants to study their use and attitudes towards digital identity over a full life span, with specific examinations of concepts such as trust, privacy, and security.

3.2 Sampling Approach

We will select a random sample of participants at an early age, and record data from the same sample once a year. The Millennium Cohort Study followed participants since they were born, while other studies such as the COSMO study recruited respondents at the age of 14 or 15. We will recruit participants at the age of 10. Several studies have shown that the vast majority of British teenagers aged 14 or more routinely use the internet (over 90%; UK Government, 2019), and are therefore likely to have some level of involvement with digital identity technologies. It is nonetheless less clear at which age individuals begin to engage with digital systems. There was a general consensus at the consultation meetings that individuals should be older than 8 and younger than 13-14 at the time of the first wave of the study - hence we take the middle-point of 10 years of age.

Participants at the methodological consultations strongly supported the idea that the sample should be composed of at least 200 participants, and ideally larger to ensure its representativeness of the target population. It was also suggested that a stratified random sample approach would be ideal to ensure that the selected sample reflects the different population groups in British society, and potentially their different level of involvement with digital technologies and digital identity. The suggested strata discussed in the meetings included differences based on country, income deprivation, ethnicity and sex. Other classifications, as well as more complex sampling designs, should be considered in the future. We thus follow a stratified random sampling design, but other similar sampling approaches could also be considered.

As an example, here we will calculate an ideal sample size based on the following subpopulations: country (England or Wales), income (highly deprived or not), ethnicity (White or BAME) and sex (female or male). Based on this, in total we have 16 subpopulations. We calculate the sample for England and Wales in this example, but the study can be expanded to Scotland and Northern Ireland as well. We first downloaded data aggregated from the UK Census 2011 for a set of demographic variables, at the level of Output Areas (website: https://www.nomisweb.co.uk/census/2011). We then generated a synthetic population database following a multivariate truncated normal and binary distribution (Demirtas et al., 2014) based on parameters recorded in the Census in each output area (for details, see Brunton-Smith et al., 2024). This allowed us to estimate the number of residents at the convergence of these different subpopulation groups. For instance, we estimate that there are in total 930,449 residents aged over 9 and under 11 in England and Wales, amongst which 92,633 are females, BAME, not deprived and English, and 1,544 are males, white, deprived and Welsh. This was calculated for each of our 16 subpopulations.

An ideal sample size for a population of 930,449 with a confidence level of 95% would be 385 participants, while a sample of 666 respondents would enable estimates with a confidence level of 99%. Attrition is nonetheless a common issue in longitudinal research designs, which should where possible be considered at the sampling design stage. Attrition refers to the decrease in sample size from one wave to another, which occurs when certain participants terminate their involvement with the project. Attrition rates from 30 to 70% are common in the literature (Gustavson et al., 2012). If we aim to compensate for a probable average 40% of attrition, for at least 10 consecutive waves, to ensure that the eleventh sample is still representative of the target population, the sample size should consist of at least 63,672 respondents in the first wave to enable estimates for waves 1 to 11 with a confidence level of 95%. During the consultation meetings, nonetheless, it was discussed that attrition tends to be less severe than 40%, and it is often more severe during the first few waves (and at the age when participants join university) than later on. Sumner (2006) notes that the average attrition rate in longitudinal studies is 17%. If we account for all this in our calculations, assuming that attrition will decrease by 5% in each wave, beginning with a 50% attrition in wave 2, an ideal sample size to enable for estimates with a confidence level of 95% at wave 11 would be 11,761 respondents. The anticipated attrition rates for waves 1 to 11 are displayed in Figure 3, and the exemplar ideal stratified sample size is detailed in Table 2. Participants will be selected randomly within each stratum.

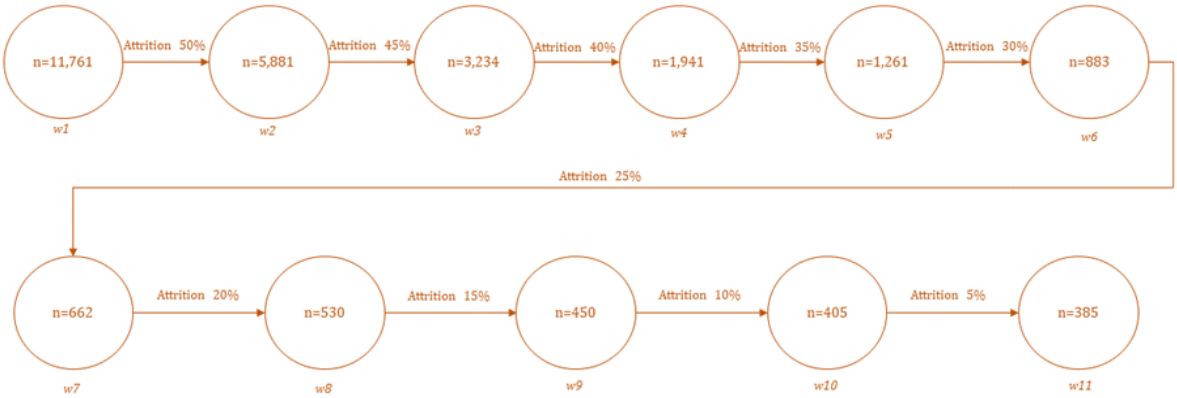

Figure 2. *Probable attrition rates and sample sizes between waves 1 and 11*

Table 2. *Stratified sample sizes for wave 1 (exemplar)*

		Male		Female	
		Highly deprived	Not highly deprived	Highly deprived	Not highly deprived
England	White	209	3,947	349	3,755
	BAME	164	1,313	249	1,171
Wales	White	20	251	32	236
	BAME	6	28	7	25

Research has further explored strategies to mitigate the risk of attrition in longitudinal studies (Fumagalli et al., 2013; Given et al., 1990), including the use of tailored invitation letters and reports for participants and their parents/guardians, the use of mixed mode surveys (see following section), and incentives in the form of monetary compensation if deemed suitable and ethically appropriate (Singer and Couper, 2008). Furthermore, missing data imputation methods are commonly used in longitudinal data analysis to compensate for missing data resulting from the attrition of participants (Twisk and de Vente, 2002). Participants in our consultation meetings also mentioned the possibility of recording data from a second and third cohort at a later stage.

Importantly, we envision DIALCS not only to record quantitative data from participants over time, but also to capture qualitative insights from smaller samples of respondents (between 15 and 20 participants) in each wave. This was recommended by participants in our consultation meetings, and would build on ongoing data collection endeavours within the UK Government. The selection of participants for the qualitative interviews may be driven by researchers' interests in particular aspects of digital identity observed in certain respondents, or interesting case studies, and hence does not necessarily need to be driven by a random selection.

3.3 Mode of Data Collection

Data will be primarily recorded through a combination of computer-assisted telephone interviewing (CATI), computer-assisted personal interviewing (CAPI), and face-to-face interviewing. Under-16 participants will first complete a set of basic questions about themselves, their family, access to internet, digital literacy, and their use and perceptions of digital identity technologies through CATI, while the larger part of the questionnaire, including detailed questions about digital identity attributes and other related items, will be completed through face-to-face interviewing at the residence of respondents (Fumagalli et al., 2013). The CATI questionnaire should not last more than 20 minutes, while the face-to-face interviewing should last between 30 and 45 minutes. For participants aged 16 or more, CATI will be replaced by CAPI for the first set of questions, while the rest of the interview would still take place in-person. We thus envision two slightly different data collection modes for under-16 and over-16 participants, as suggested by attendants to our consultation meetings. This aims to reduce the risk of attrition. Data collection will be undertaken by a company or university department specialised in survey panel services. Qualitative interviews will take place through face-to-face interviewing.

Interestingly, some participants in our consultation meetings discussed the possibility of linking data from participants to their social media data, in order to obtain in-depth first-hand understanding of their use of digital technologies and digital identity. The research team considered this possibility and agreed that the ethical concerns associated with linking social media data to participants' survey responses outweigh the anticipated benefits of recording and linking this data, and hence we do not envision recording extra information from online sources for participants in the study.

3.4 Questionnaire Design and Validation

We will design a questionnaire to measure all relevant aspects of perceptions, attitudes, and behaviours towards digital identity. The questionnaire will be designed alongside relevant stakeholders working in public sector organisations (e.g., Department for Digital, Culture, Media

and Sport, International Organization for Standardization, Home Office, Department for Science, Innovation and Technology, police forces), to ensure its content is relevant both to researchers and practitioners and policy makers. The questionnaire will be designed in four stages:

- first, a list of key constructs will be developed (see first list below);
- second, validated tools to measure such constructs will be searched in existing literature and adapted to the UK reality where needed;
    - where needed, new items will be designed;
- third, a pilot study of the first wave with 59 participants (Viechtbauer et al., 2015) will be used to assess the questionnaire and, where needed, validate new items (tests of measurement invariance across cultures and population domains will be undertaken); and
- fourth, finalise the final questionnaire and adapt it to CATI, CAPI and face-to-face interviewing modes.

We envision the questionnaire to be divided in two main forms: the 'general screening form' and the 'digital identity form(s)'. The 'general screening form' will include questions about the use of digital devices, digital access to various platforms and sites, perceptions about digital technologies, perceptions about digital identity technologies, digital literacy, parental supervision and control, experiences with digital technologies, individual traits, and demographic and social characteristics (see Table 3). Importantly, for every digital device and platform identified, the questionnaire will ask respondents if the attributes used to access and navigate/use the device or platform perfectly align with their formally assigned identity (i.e., same demographic and social characteristics as 'offline' self). Every time a respondent acknowledges a different, even a slightly different, set of attributes used in a specific device or platform, the questionnaire will activate a new 'digital identity form' to record further information about each 'digital identity'. The maximum number of 'digital identity forms' per respondent will be capped at 7 to reduce the risk of fatigue and non-response bias.

Table 3. *Topics and items included in 'general screening form'*

Topic	Items
Digital devices	Access to, and number of hours spent in, various digital devices (e.g., smart phone, smart watch, laptop, personal computer, tablet, gaming consoles, smart TV, smart home devices). Include follow-up questions for each device (e.g., single or shared use (and if shared, with whom), connected to the internet or not).
Digital access	Access to, and number of hours spent in, various digital platforms and sites (e.g., social media sites, online forums, gaming platforms, digital newspapers, dark nets, government digital sites, education platforms, e-betting sites, dating apps and websites, e-shopping sites, food delivery services, online banking apps and websites, digital currency services, streaming services).
Perceptions about digital technologies	For every digital device and platform identified above, include follow-up questions about perceptions of security, privacy, and trust. Also include a general question of perceived security, privacy, and trust in digital technologies.
Perceptions about digital identity technologies	General awareness of digital identity technologies. Perceived security, privacy, and trust in more or less advanced digital identity technologies (e.g., ID cards, digital signatures, face recognition, fingerprint scanning, password verification, username, decentralised identifiers (with a simple explanation), encrypted keys and identifiers). Directly define 'digital identity' (open answer).
Perceived digital literacy	Perceived ability to find information on digital platforms. Perceived ability to evaluate information on digital platforms. Perceived ability to communicate information on digital platforms.
Experiences with digital technologies	Victimisation and self-reported deviant behaviour through digital environments (e.g., cyber bullying, ID fraud, online shopping fraud, romance fraud, computer virus, hacking, digital access and sharing of copyright-protected material, sexting, online pornography, revenge porn, e-purchasing of illegal items (drugs, weapons, malware, personal data); Buil-Gil et al., 2024). For every victimisation identified, include follow-up questions about harm and help-seeking responses.
Digital self-protection	Use of digital security technologies (e.g., antivirus software, encryption, pseudonymisation) and self-protection behaviour (e.g., do not talk to strangers online, do not share images online, do not engage in online shopping).
Parental supervision and control	Parental control and supervision over respondent's everyday activities, online and offline (e.g., parents know where I am when I am with friends, parents control the number of hours I spend connected to the internet, parents control that I complete my homework, parents ask me to at home at a certain time).
Individual traits	Self-control, sociability, risk-taking tendencies (online and offline), body type (hair colour, eye colour, height, weight), self-esteem, self-perceived attractiveness.
Demographic, physical and social characteristics	Age, sex, gender identity, ethnicity, income, area of residence, country of birth, nationality, primary language, education level, academic performance, disabilities, support networks (e.g., family, school, friends).
Digital identity screening	For every digital device and platform identified above, ask if the attributes used to access, and navigate/use, the device or platform, perfectly align with the formally assigned identity of the respondent (i.e., same demographic, physical and social characteristics as 'offline' self, including profile picture or avatar).

For every new 'digital identity' identified in the 'general screening form', a 'digital identity form' will ask participants follow-up questions about a set of attributes. The maximum number of 'digital identity forms' per respondent will be capped at 7.

'Digital identity forms' will include detailed questions about the attributes of each identified 'digital identity' (e.g., expressed age, sex, gender identity, ethnicity, income, body type, area of residence, country of birth, primary language, nationality, education level, academic performance and support networks (if any)), use of photos and avatars, digital platforms where this identity is adopted, number of hours spent with this 'identity', financial transactions (if any), number of people the identity links with (and type of links), whether the identity is shared with someone else, self-perceived attractiveness and 'personality' of digital identity, and general uses of the digital identity (e.g., professional, recreational, health-related, deviant/criminal, sexual).

3.5 Data Dissemination, Access and Use

We aim to disseminate all data recorded by DIALCS, including both the quantitative and qualitative datasets, through the repository UK Data Service (https://ukdataservice.ac.uk/). Quantitative data will be publicly available for registered users, while qualitative data will be considered 'secure' and will only be available upon request for researchers accredited by the Office for National Statistics or the UK Data Service.

Sharing the data openly will enable researchers and students based across the social, technological and behavioural sciences to access this data and further contribute to digital identity research as well as the development and evolution of policy and architectural standardisation. In short, DIALCS can set the foundations for a new multidisciplinary body of research on life-course digital identity, while also providing key data for researchers in related fields. This will impact how policy-makers can develop ongoing digital identity policy frameworks, and will impact how technology developers construct processes and protocols in cyberspace. Learning how technological constructs impact online behaviours will also allow society a clear and transparent manner for assessing how DI technology impacts on who we are, who we want to be, and who we 'ought' to be.

3.6 Ethical Considerations

This study, while promising in its scope and potential for insights, raises some ethical considerations that demand our attention.

Informed Consent and Privacy: The cornerstone of ethical research involving the collection of personal data is informed consent. In the context of DIALCS, participants should continually provide informed consent as they transition from late childhood to adulthood. Given the sensitive nature of digital identity, it is of paramount importance to ensure that participants are fully informed about how their data is used, who has access to it, and the measures in place to anonymise and safeguard their information.

Longitudinal Research and Attrition: It is incumbent upon researchers to make diligent efforts to mitigate attrition and consider the ethical aspects related to participant retention and

compensation. Researchers must maintain transparency regarding potential attrition rates and their impact on the validity of the study.

Data Collection Modes: The study encompasses various data collection methods, including face-to-face interviews, CATI and CAPI. Researchers should ensure that data collection methods do not cause distress or discomfort to the participants.

Questionnaire Design and Validation: The DIALCS questionnaire should undergo rigorous ethical review. Questions related to personal characteristics, digital identities, and sensitive experiences must be handled with care to prevent psychological harm. Measures will be in place to ensure that respondents comprehend the questions posed and support is provided in case of need.

Data Dissemination: While open access to data is laudable for research and policy, the release of both quantitative and qualitative data must be carefully controlled to protect participants' privacy and anonymity. Implementation of ethical safeguards, including data anonymisation and secure access mechanisms, is important. Each participant will be assigned a unique code and pseudonymisation processes will be followed to ensure anonymity. Data will be stored securely within university protected servers.

Community and Stakeholder Involvement: In line with ethical research principles, it is crucial to involve relevant stakeholders in the research process. Engaging with these groups can help ensure that the study respects local cultural norms and can identify and address ethical concerns that may not be immediately apparent.

In summary, the DIALCS study presents several ethical considerations that encompass informed consent, privacy, data collection, sampling, and data dissemination. These ethical implications should be addressed comprehensively to ensure that the study is conducted with the utmost respect for the rights and well-being of the participants. Regular ethical reviews and consultations with experts in ethical research can help ensure that the study remains ethically sound throughout its duration.

4. Way Forward

DIALCS will produce a novel research framework to allow researchers, industry practitioners, and policymakers to better understand how the development of identity technology impacts one's self-presentation online. The research framework will also devise measurement indicators for the technologies impact on peoples' lives, such as their involvement in professional, personal and health-related activities. The longitudinal study will be conducted in parallel with the deployment and evolution of both digital identity technology and policy, positioning itself as a valuable validation exercise, to ensure that positive and negative impacts of the technology are measurable, verifiable and, ultimately, manifesting as expected. The study will also act as a critical trust and transparency tool, informing stakeholders of how the technologies deployment is being received by civil society. In this way, DIALCS provides a step change to how we think of digital identity and what technology means for the real world.

The impact of DIALCS will have significant implications for the scholarly understanding of digital identity, as well as for industry and policy. From an academic perspective, the collection of longitudinal data regarding perceptions, attitudes, and behaviours related to digital identity, both in quantitative and qualitative forms, will empower researchers to address pivotal questions. These questions encompass inquiries such as "What influences the development of one's online 'self' identity over their lifetime?", "How do individuals shape their 'self' identity in diverse digital environments?", "What factors drive individuals' decisions to engage with specific digital identity technologies while eschewing others?", "What are people's views on security and privacy concerning digital identity technologies, and how do these views impact the formation of digital identities?", and "What indicators reflect the use and experience of digital identity (e.g., interactions per pseudonym, pseudonym duration, etc.)?"

The implementation of DIALCS in Phase 2 will depend upon access to further resources and funding.

References

Anaraky, R. G., Byrne, K. A., Wisniewski, P. J., Page, X., & Knijnenburg, B. P. (2021). To Disclose or Not to Disclose: Examining the Privacy Decision-Making Processes of Older vs. Younger Adults. In Y. Kitamura, A. Quigley, K. Isbister, P. Bjorn & S. Drucker (Eds.), *Proceedings of the Conference on Human Factors in Computing Systems*. Association for Computing Machinery.

Arkin, R. M. (1981). Self-Presentation Style. In J. T. Tedeschi (Ed.), *Impression Management Theory and Social Psychological Research* (pp. 311-333). New York: Academic Press.

Baumeister, R. F. (1999). The Nature and Structure of the Self: An Overview. In R. Baumeister (Ed.), *The Self in Social Psychology* (pp. 1-20). Philadelphia: Psychology Press.

Baumeister, R. F. (2011). Self and Identity: A Brief Overview of What They Are, What They Do, and How They Work. *Annals of the New York Academy of Sciences*, 1234(1), 48-55.

Belen-Saglam, R., Altuncu, E., Lu, Y., & Li, S. (2023). A Systematic Literature Review of the Tension Between the GDPR and Public Blockchain Systems. *Blockchain: Research and Applications*, 4(2), 100129.

Brandtzæg, P. B., Lüders, M., & Skjetne, J. H. (2011). "Too Many Facebook 'Friends'? Content Sharing and Sociability Versus the Need for Privacy in Social Network Sites": Corrigenda. *International Journal of Human-Computer Interaction*, 27(1), 106.

Brunton-Smith, I., Buil-Gil, D., Pina-Sánchez, J., Cernat, A., & Moretti, A. (2024). Using Synthetic Crime Data to Understand Patterns of Police Under-Counting at the Local Level. In L. Huey & D. Buil-Gil (Eds.), *The Crime Data Handbook*. Bristol: Policy Press.

Buil-Gil, D., Trajtenberg, N., & Aebi, M. F. (2024). Measuring Cybercrime and Cyberdeviance in Surveys. In *Routledge Handbook of Online Deviance*. Routledge.

Burke, C., & Moltorisova, A. (2017). What Does It Matter Who is Browsing. *Journal of Intellectual Property, Information Technology and Electronic Commerce Law*, 8, 238.

Carver, C. S., & Scheier, M. F. (1981). *Attention and Self-Regulation: A Control Theory Approach to Human Behavior*. New York: Springer.

Chen, B., & Marcus, J. (2012). Students' self-presentation on Facebook: An examination of personality and self-construal factors. *Computers in Human Behavior*, 28(6), 2091-2099

De Hert, P. (2008). Identity Management of e-ID, Privacy and Security in Europe. A Human Rights View. *Information Security Technical Report*, 13(2), 71-75.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. New York: Plenum Press.

Demirtas, H., Amatya, A., & Doganay, B. (2014). BinNor: An R Package for Concurrent Generation of Binary and Normal Data. *Communications in Statistics - Simulation and Computation*, 43(3), 569-579.

Durrant, A., Trujillo-Pisanty, D., Moncur, W., & Orzech, K. (2015). *Charting the Digital Lifespan: Picture Book*. University of Newcastle.

Emanuel, L., Neil, G. J., Bevan, C., Fraser, D. S., Stevenage, S. V., Whitty, M. T., & Jamison-Powell, S. (2014). Who Am I? Representing the Self Offline and in Different Online Contexts. *Computers in Human Behavior,* 41, 146–152.

Friedenberg, J. (2020). *The Future of the Self: An Interdisciplinary Approach to Personhood and Identity in the Digital Age*. University of California Press.

Fumagalli, L., Laurie, H., & Lynn, P. (2013). Experiments with Methods to Reduce Attrition in Longitudinal Surveys. *Journal of the Royal Statistical Society Series A*, 176(2), 499-519.

Gaggioli, A. (2017). Artificial Intelligence: The Future of Cybertherapy? *Cyberpsychology, Behavior, and Social Networking*, 20(6), 402-403.

Gaither, S. E. (2018). Belonging to Multiple Groups: Pushing Identity Research Beyond Binary Thinking. *Self & Identity*, 17, 443-454.

Gibbs, J. L., Ellison, N. B., & Heino, R. D. (2006). Self-Presentation in Online Personals. *Communication Research*, 33(2), 152-177.

Given, B. A., Faan, L. J., Collins, C., & Given, C. W. (1990). Strategies to Minimize Attrition In Longitudinal Studies. *Nursing Research*, 39(3), 184-187.

Goffman, E. (1959). *The Presentation of Self in Everyday Life.* Doubleday.

Goodell, G., & Aste, T. (2019). A Decentralized Digital Identity Architecture. *Frontiers in Blockchain*, 2, 17.

Gstrein, O. J., & Kochenov, D. (2020). Digital Identity and Distributed Ledger Technology: Paving the Way to a Neo-Feudal Brave New World? *Frontier in Blockchain*, 3(10).

Gustavson, K., von Soest, T., Karevold, E., & Røysamb, E. (2012). Attrition and Generalizability in Longitudinal Studies: Findings from a 15-year Population-Based Study and a Monte Carlo Simulation Study. *BMC Public Health*, 12, 918.

Habib, H., Shah, N., & Vaish, R. (2019). Impact of Contextual Factors on Snapchat Public Sharing. In S. Brewster, G. Fitzpatrick, A. Cox & V. Kostakos (Eds.), *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-13). Association for Computing Machinery.

Hew, K. F. (2011). Students' and Teachers' Use of Facebook. *Computers in Human Behavior,* 27(2).

Hollenbaugh, E. E., & Ferris, A. L. (2015). Predictors of Honesty, Intent, and Valence of Facebook Self Disclosure. *Computers in Human Behavior*, 50, 456-464.

Higgins, E. T. (1987). *Self-Discrepancy: A Theory Relating Self and Affect.* Psychological Review, 94, 319-340.

Huang, J., Kumar, S., & Hu, C. A. (2021). Literature Review of Online Identity Reconstruction. *Frontiers in Psychology*, 12, 696552.

Hussain, Z., Griffiths, M. D., & Baguley, T. (2012). Online Gaming Addiction: Classification, Prediction and Associated Risk Factors. *Addiction Research & Theory,* 20, 359-371.

Jin, S. A. (2012). The Virtual Malleable Self and the Virtual Identity Discrepancy Model: Investigative Frameworks for Virtual Possible Selves and Others in Avatar-Based Identity Construction and Social Interaction. *Computers in Human Behavior,* 28(6), 2160-2168.

Katz, E., Blumler, J. G., & Gurevitch, M. (1973). Uses and Gratifications Research. *Public Opinion Quarterly*, 37(4), 509–523.

Kurek, A., Jose, P. E., & Stuart, J. (2017). Discovering Unique Profiles of Adolescent Information and Communication Technology (ICT) Use: Are ICT Use Preferences Associated with Identity and Behaviour Development? *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 11(4), 3.

Lee, D. B., Goede, J., & Shryock, R. (2010). Clicking for Friendship: Social Network Sites and the Medium of Personhood. *MedieKultur. Journal of Media and Communication Research*, 26 (49): 137–150.

Li, S., Zheng, J., & Chiang, F. (2022). Examining the Effects of Digital Devices on Students' Learning Performance and Motivation in an Enhanced One-to-One Environment: A Longitudinal Perspective. *Technology, Pedagogy and Education*, 31(1), 1-13.

Limniou, M., Lu, Y., Renwick, R., Buil-Gil, D., & Zeng, Y. (2023). *Identity in the Digital Age: An Analysis of Terminological Disparities in 'Digital Self' Theories and Digital Identity Frameworks*. Article under review.

Masur, P. K., Bazarova, N. N., & DiFranzo, D. (2023). The Impact of What Others Do, Approve Of, and Expect You to Do: An In-Depth Analysis of Social Norms and Self-Disclosure on Social Media. *Social Media + Society*, 9(1).

Meeus, W. (2011). The Study of Adolescent Identity Formation 2000–2010: A Review of Longitudinal Research. *Journal of Research in Adolescence*, 21(1), 75-94.

Mitrushchenkova, A. N. (2022). Personal Identity in the Metaverse: Challenges and Risks. *Kutafin Law Review*, 9(4), 793-817.

Newman, J. C., Des Jarlais, D. C., Turner, C. F., Gribble, J., Cooley, P., & Paone, D. (2002). The Differential Effects of Face-to-Face and Computer Interview Modes*. American Journal of Public Health*, 92(2), 294–297.

Orzech, K., Moncur, W., Durrant, A., & Trujillo-Pisanty. (2016). Opportunities and Challenges of the Digital Lifespan: Views of Service Providers and Citizens in the UK. *Information, Communication and Society*, 21, 14-29.

Prescott, T. J., & Robillard, J. M. (2020). Are Friends Electric? The Benefits and Risks of Human-Robot Relationships. *iScience*, 24(1), 101993.

Quan-Haase, A., & Elueze, I. (2018). Revisiting the Privacy Paradox: Concerns and Protection Strategies in the Social Media Experiences of Older Adults. In A. Gruzd, J. Jacobson, P. Mai, J. Hemsley, K. H. Kwon, R. Vatrapu, A. Quan-Haase, L. Sloan &. J. Hodson (Eds.), *Proceedings of the 9th International Conference on Social Media and Society* (pp. 150-159). Association for Computing Machinery.

Rogers, C. R. (1959). A Theory of Therapy, Personality, and Interpersonal Relationships as Developed in The Client-Centered Framework. In S. Koch (Ed.), *Psychology: A Story of a Science* (pp. 184-256). New York: McGraw-Hill.

Rogers, C. R. (1961). *On Becoming a Person: A Psychotherapist's View of Psychotherapy*. New York: Houghton Mifflin.

Schreier, N., Renwick, R., & Ehrke-Rabel, T. (2021). The Digital Avatar on a Blockchain: E-Identity, Anonymity and Human Dignity. *Austrian Law Journal*, 2021, 202-218.

Sibilla, F., & Mancini, T. (2018). I Am (Not) My Avatar: A Review of the User-Avatar Relationships in Massively Multiplayer Online Worlds. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 12(3), 4.

Singer, E., & Couper, M. P. (2008). Do Incentives Exert Undue Influence on Survey Participation? Experimental Evidence. *Journal of Empirical Research on Human Research Ethics*, 3(3), 49–56.

Stevenage, S., Black, S., Creese, S., Gust, R., Saxby, S., Love, O., Fraser, D., Whitty, M., Bevan, C., & Emanuel, L. (2013). *SuperIdentity: Linking Online and Offline Identities: Project Annual Report (Year Two): 2013*. EPSRC.

Sumner, M. (2006). Attrition. In M. Sumner (Ed.), *The SAGE Dictionary of Social Research Methods*. SAGE.

Twisk, J., & de Vente, W. (2002). Attrition in Longitudinal Studies: How to Deal with Missing Data. *Journal of Clinical Epidemiology*, 55(4), 329-337.

UK Government (2019). *Internet Use.* Retrieved from https://www.ethnicity-facts-figures.service.gov.uk/culture-and-community/digital/internet-use/latest (Last access 23 October 2023).

Van Dijck, J. (2013). 'You Have One Identity': Performing the Self on Facebook and LinkedIn. *Media, Culture & Society*, 35(2), 199-215.

Van Dijck, J., & Jacobs, B. (2020). Electronic Identity Services as Sociotechnical and Political-Economic Constructs. *New Media & Society*, 22(5), 896-914.

Viechtbauer, W., Smits, L., Kotz, D., Bude, L., Spigt, M., Serroyen, J., & Crutzen, R. (2015). A Simple Formula for the Calculation of Sample Size in Pilot Studies. *Journal of Clinical Epidemiology*, 68(11), 1375-1379.

Viik, T. (2020). Falling in Love with Robots: A Phenomenological Study of Experiencing Technological Alterities. *Paladyn, Journal of Behavioral Robotics*, 11(1), 52-65.

Walther, J. B. (1992). Interpersonal Effects in Computer-Mediated Interaction: A Relational Perspective. *Communication Research*, 19(1), 52–90.

Walther, J. B. (1996). Computer-Mediated Communication: Impersonal, Interpersonal, and Hyperpersonal Interaction. *Communication Research*, 23(1), 3–43.

Weir, K. F., & Jose, P. E. (2010). The Perception of False Self Scale for Adolescents: Reliability, Validity, and Longitudinal Relationships with Depressive and Anxious Symptoms. *British Journal of Developmental Psychology*, 28, 393-411