N/UX

Designing the Non-User Experience

Timothy Quinn

DARKDATA project

N/UX

Designing the Non-User Experience

Timothy Quinn
Dark Data Project
darkdataproject.org

v 2023.02

Introduction

Although now all but ubiquitous, user experience design, or UX, is a relatively young discipline which popularly traces its roots to the period of innovation in the 1980s and 1990s during which the graphical user interface accelerated the adoption of personal computers, and designers like Don Norman and David Kelley extended user-centric industrial design concepts like affordances and conceptual mapping into the ephemeral world of software. With any new discipline, initial use cases tend to be narrow (e.g. the consumer operating system) and then expand over time across industries with the realization that tools and methodologies developed for one class of problem may apply to many others. Today, a job search for "UX" is as likely to return results from pharmaceutical companies and municipal governments as from design or technology firms, and applications of UX design continue to expand exponentially.

As diverse as UX use cases have become, there remain countless pockets of unexplored applicability for which UX is largely untested. The design and management of non-user, or "third party," experience is one of these.

"Third party" or "end recipient" individuals are introduced in greater depth in a previous white paper, Anthroencryption: Strategies for Protecting the World's Most Vulnerable People[1], which focuses primarily on issues of data protection and shareability. This paper focuses instead on the applicability of UX principles and methodologies to the problem of managing individuals with little or no awareness of, or agency over, their data in any given system, arguing that by refocusing broadly on "people" rather than just "users" a UX-based approach can be brought to bear on a problem that is too often dismissed as solely one of security.

What is a Non-User?

Refugees who cross borders in their flight from natural disasters, economic crises, mass atrocities, ethnic or religious persecution, or other forms of existential threat often wind up as data in the support systems of other governments and international aid organizations, and may do so under levels of duress which negate any awareness, consent or agency over their data. They have become non-users of those databases which store their most personal and endemic data. Likewise, people who utilize the support infrastructure within their own countries, such as

[1] Dark Data Project, 2022

food banks and homeless shelters, are often compelled to provide identifying data (sometimes to prevent duplication of services, sometimes for statistical or demographic purposes) while having little control over that data after the moment of intake. Renting an apartment or applying for a car loan usually involves a credit check with a large consumer data broker, where again many will have little awareness of the data compiled about them (a problem which is exacerbated by historical problems among data brokers of security, privacy and accuracy).

If you've been the witness to a crime, provided confidential information to law enforcement or the media, or engaged in any activity that would cause your elected or unelected officials concern, if perhaps you've traveled to a sanctioned country or participated in a protest or even visited a website your government has blocked, you have similarly become a non-user: your identity has been added to a database with or without your consent and you are now known by the data which is harvested about you and over which you have little or no control.

It's perhaps natural outside of autocratic regimes to wonder how a person can provide data without being aware of the retention of that data, and therefore exercise agency over it. How can anyone be so naive? Part of the problem is the ease with which we confuse data "in transit" and data "at rest," terms which will be familiar to anyone who's had to secure networked data; data in transit is inherently transactional, moving from point A to point B, whereas data at rest is data stored for later use. When your passport is scanned at a border crossing, you assume your identity is transiting a system that verifies you are who you claim to be, but you might not assume that the record of that transit winds up resting in a database administered by, say, your country's domestic intelligence service. Similarly, you may have a level of awareness that advertisers are tailoring the ads you see online based on your social media activity, but you may not be aware that this personalization can be based on the retention of audio recordings of private offline conversations held within earshot of your mobile phone.

And then, of course, there's the other part of the problem, which is that lack of awareness or agency over one's data has a disproportionate impact on vulnerable people, specifically marginalized or disenfranchised populations and those in crisis. Those who own their homes or

don't require multiple lines of credit to manage their monthly expenses are inevitably less concerned about their credit scores, and thus less inclined to advocate for redress and transparency, than those whose livelihoods are directly impacted by the risk of not having a place to live or a way to get to work. For those fleeing war and persecution, the disparity is even more acute: the people with the least authority over their data are those who suffer the greatest impact when that data is inaccurate, mishandled or used to their detriment.

Non-user data is a subset of a much broader problem, which is the problem of dark data: data which is inaccessible and unaccountable, which can be unstructured and inaccurate, untranslated or untranslatable, used for purposes that are illicit or inappropriate, or is in various other ways outside the purview of those from whom data has been collected. We've become accustomed to the idea that we live in a data-saturated world and that our best decisions are evidence-driven, drawn from the vast ecosystem of data providers that empower our global economy; as a result, many of us struggle to balance expectations of privacy against the perception that every waking moment of our lives simply creates more data for others to monetize. What's missing from this paradigm is both the data we're unaware of and the data we presume exists but doesn't, such as the dauntingly high number of births and deaths around the world that go unrecorded in any government ledger. Although the problem of dark data is greater than just identity data, and the problem of non-user data is more than just an issue of accuracy and agency, the non-user experience is a unique opportunity for collaboration that exists at the intersection of several disciplines, where designers have as much a role to play as software engineers and CIOs.

The Designer, the Data Scientist and the Security Analyst

Onboarding non-user data is accomplished in any number of ways, but can be broadly categorized as being first-party (a person provides data, assuming its use in transit but not at rest), second-party (an interviewer solicits data directly from a person) or third-party (data is acquired from an intermediary, such as a data provider or data broker). The intake or ingestion process results in data being retained in some form (call it a database, although the actual type of data and storage mechanisms will vary). The retained data is then typically accessed in some manner for purposes of authentication or analysis.

This basic workflow implies the participation of several different roles, and reveals a much more complex interaction of competencies and biases than the simplistic model of a faceless data

harvester with a single Berserker-like goal of boundless information collection. There is, in some capacity, a design process at work, which requires the skills of an information designer, both to ingest data and to make it useful for analysis. There is also clearly a data architect or scientist who must decide how data is to be stored for optimal performance, and a security analyst or engineer whose job is to ensure that the ingestion, storage and delivery of data is done safely and in a manner that mitigates the risk of compromise.

If it comes as a surprise that a designer is an essential component in the collection and handling of non-user data, it will also come as a surprise that the non-user experience can be as design-intensive an exercise as architecting a user-driven experience, and that many of the proven design principles, methodologies and tools of UX are similarly valid within the context of N/UX. In addition, there are several design principles which are unique to N/UX, and these will be the primary focus of the rest of this paper.

Useful N/UX Design Principles

The following principles address opportunities at the point of data intake, as well as best practices for the analysis and use of data at rest. Some are design-specific interpretations of security strategies discussed in greater depth in Anthroencryption: Strategies for Protecting the World's Most Vulnerable People[2].

Help non-users become users

A fundamental assumption of all privacy legislation is that people should have agency over their data. As such, one of a designer's top priorities should be facilitating the migration of non-users into users. This is obviously dependent on non-users being aware of the retention of their data in a given system, and is complicated by the reluctance of non-users to provide further data to verify their ownership over their data.

A good design pattern to follow is to provide a prominent call to action in front of any repository of non-user data, and then query for the minimal amount of data necessary to validate the inquiring party's identity. Once the identity of the inquiring party has been validated, if no matching non-user account has been found, a "no results" finding should be communicated and no non-user data should be created as a result of this interaction. If a matching non-user

[2] Dark Data Project, 2022

account has been found, the inquiring party should be granted the authority to administer that data, thus becoming a user.

In general, credit agencies follow this methodology, albeit with occasionally overzealous collection of verification data, and it's unclear which of them may use the interaction as an opportunity to create a new non-user account. (Does a person who's concerned about their credit warrant a lower credit score? Most credit agencies say no, although validations performed by third parties does appear to result in a perception of heightened risk.)

Law enforcement follows a similar methodology: the United States Department of Justice, for instance, allows people to file FOIA (Freedom of Information Act) requests about themselves by submitting notarized proof of identity, although it's similarly difficult to state with certainty that any given law enforcement entity won't interpret a request for information on a non-existent non-user as an opportunity to create such a record.

In all cases, it is incumbent upon the designer to architect workflows which simplify the process of transitioning non-users into users, and minimize the negative impact on people who are not non-users.

Salt it, hash it, then validate it

When validating a user's identity, it's common to ask for sensitive, and frequently unchangeable, personal data such as date and place of birth, government ID number, email address, telephone number, or mother's maiden name. The purpose of taking this information is to match it against known data on non-users and find a match.

A prudent design practise when matching identity data is to hash it first (i.e. encrypt it), and then compare hashes instead of cleartext (i.e. unencrypted data). This keeps sensitive data from transiting any systems or being stored accidentally or intentionally. This is therefore currently the recommended approach for password logins (although it's unfortunately still possible to find organizations that store passwords in cleartext or create hashes with insecure algorithms).

When hashing identity data, it's also important to "salt" it, which involves adding an environment- or context-specific amount of additional data prior to hashing. This ensures that any hashes are unique to a given environment or context, and in the event those hashes are retained, they won't match other hashes in the wild. Comparing hashes across domains is a

form of correlation inference that can leverage data breached from one source to unencrypt data from another source.

Aggregate; don't disambiguate

A core principle of N/UX which is borrowed from a similar principle in UX and data security is to architect for the least possible implementation: design for need to know. Non-user data may be retained for a variety of reasons, not all of which require individualization, and may be viewed by a variety of people operating in a variety of roles, not all of whom again may require data at the level of the individual.

Where possible, aggregate, and do so safely. Safe aggregation means displaying resultsets that are large enough to obfuscate unique indicators of individuals. Data reduction through filtering is one means of accidental disclosure; if a dataset is based on a small number of respondents, and that data can be further winnowed down by attributes like location or demographics, it's not impossible to wind up with a dataset representing a single individual.

Disambiguation should be particularly avoided. If a designer has already had the forethought to responsibly deindividualize data at rest, data analysts working with that data shouldn't employ data mining techniques designed to disambiguate that data, not only because it violates the ethical principle which informed encryption in the first place, but because, like reverse engineering, disambiguation often generates results which may resemble but not accurately match the source. An example of faulty disambiguation is the mining of consumer data and social media activity by political parties to categorize voters into overly specific and generally ridiculous categories that voters themselves would find of condescending and inaccurate (e.g. "soccer moms" and "NASCAR dads").

Don't do predictive collection

There is a depressingly common tendency when designing data collection methodologies to mitigate unknown future needs by asking for an excess of data, and it's almost always a bad idea. Data which doesn't have an immediate and impactful need is little more than an escalation in risk for the non-users whose data has been captured, a risk which has been demonstrated again and again through high-profile data breaches.

As above, designers should adhere to the principle of least possible implementation: architect intake workflows that capture as little data as possible, and no data at all which serves ambiguous or unknown purposes.

A name is not a person

If you've ever had the opportunity to see a government watchlist, or the misfortune to be on one, you'll be surprised to note how low the bar is to be a watchlisted non-user. A large percentage of the watchlist data stored about citizens of high-income countries is of problematic accuracy and origin, and the data on citizens of low-income countries is often devoid of unambiguous datapoints such as government ID numbers and dates of birth. The only thing that most watchlists seem to universally rely upon is a name, and a name is not a person.

Certain watchlists used for tracking hard-to-find individuals like terrorists and money launderers sometimes have little more than a name, which is especially problematic in countries where surnames are less varied, such as Vietnam where a third of the population has the surname Nguyen and 90% of the population shares one of 14 surnames, or China, where surnames are often changed to accommodate keyboards with limited character sets and 1.3 billion people therefore share 6,000 surnames.


NOFLY.csv

```
SID,CLEARED,LASTNAME,FIRSTNAME,MIDDLENAME,TYPE,DOB,POB,CITIZENSHIP,PASSPORT/IDNUMBER,MISC
1,,MUSSADDI                                  /22/1953,,,,
2,,IAMAOUI,
6,,HYMDAN,E
16,,AL HILA                                  /1949,,,,
21,,OMAR,MC
24,,ABBAS,M                                  ,,
32,,BEN JAE
40,,SAYED,S                                  ,,,
46,,HASAN,H
73,,HAMMAMI                                  1960,,,,
113,,GOPAL                                   ,,,
115,,VERMA,
117,,VARMA,                                  ,,,
119,,VARMA,                                  ,,,
121,,KALR,F                                  ,,
124,,RAJESH                                  ,,,
```

In 2019, poor online security by US regional carrier CommuteAir resulted in the disclosure of the TSA's notorious "no fly" list, which validated the concerns of civil libertarians that sparse and inaccurate data was being used to target travelers of primarily Middle Eastern ethnicity, including children as young as four years old, without consideration of accuracy, transparency or redress. Over the past few decades of centralized "no fly" watchlists, an endless stream of individuals has complained publicly about being mistaken for watchlisted terror suspects based solely on their names, some going so far as to legally change their birth names to avoid endless security screenings and missed flights.

Designers should therefore ensure when designing interfaces that a deficit of data is made obvious, even if it goes against the designer's instinct to hide empty fields and columns for visual simplicity.

Asked and answered

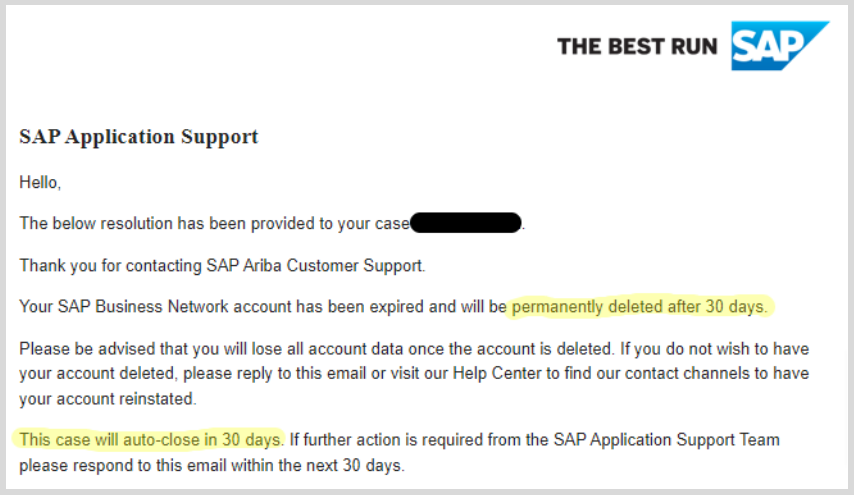
Despite the interventions of large international aid organizations like the United Nations and the Red Cross, support for refugees in transit is often a distributed effort shared by several organizations, and can result in victims of violence, persecution and other crises being forced to recount traumatizing testimony multiple times to multiple entities. This is in part a logistics and coordination problem, and in part a privacy problem, since many organizations are ill-equipped and disinclined to share sensitive victim data. Aside from compelling victims to relive horrifying experiences, a lack of data sharing also inhibits mechanisms for transnational justice.

Designers should advocate and design for the safe sharing of traumatic testimony between supporting organizations and the minimization of bureaucratic hurdles during the data intake of non-users. Leverage the work that's already been done to remove unnecessary burdens.

Promote perishability

Despite legislation in various countries to protect the rights of users and non-users to have their data deleted, the reality of data deletion is unnecessarily daunting for everyone. Few organizations provide users with the means to self-delete their data, and obviously no accommodations are made for the self-deletion of non-user data because its in no organization's interest to do so.

Designers should create workflows that allow for non-users who validate their identity to delete data stored about themselves, without the need for excessive additional data intake, and without the need for human intercession, which generally acts as a



THE BEST RUN SAP

SAP Application Support

Hello,

The below resolution has been provided to your case [redacted].

Thank you for contacting SAP Ariba Customer Support.

Your SAP Business Network account has been expired and will be permanently deleted after 30 days.

Please be advised that you will lose all account data once the account is deleted. If you do not wish to have your account deleted, please reply to this email or visit our Help Center to find our contact channels to have your account reinstated.

This case will auto-close in 30 days. If further action is required from the SAP Application Support Team please respond to this email within the next 30 days.

deliberate deterrent to account deletion.

An illustrative example of a deeply flawed design pattern is the process by which users currently delete themselves from SAP's ecosystem[3]. Users must validate themselves through login to create a support ticket, which, once processed by a human being, results in a 30-day wait for automated deletion. The support ticket, however, also expires in 30 days, which means that if the user's data is not properly deleted from the numerous internal databases maintained by SAP, the support ticket has been closed by the time the user realizes this, and the process must begin anew.

Consent under duress is gunboat diplomacy

In 1842, China signed the Treaty of Nanking with Great Britain, ending what would become known as the First Opium War, opening up trade, imposing financial indemnities, and ceding Hong Kong to the British, among other concessions. It is now historically known as the first of the Unequal Treaties because the agreement was signed aboard a British gunboat, one of many naval vessels strategically placed to ensure compliance. This was neither the first nor last example of what is commonly called gunboat diplomacy: the practice of negotiating through the imbalance of power.

Many essential services for vulnerable people require a form of consent that is compromised both by the age of some non-users and by the circumstances under which consent is provided. It's impractical and unfair to expect refugees, asylum seekers and victims of violence or trafficking to understand the implications of, and consent to, data collection as a prerequisite to intervention, particularly if the applicant isn't fluent in the language in which consent is sought, and particularly when the applicant is a minor.

Unfortunately, most forms of consent are pro forma: there is no option to request intervention without explicitly or implicity providing consent. Canada's "Basis of Claim" form[4], which is the means by which refugees apply for the protection of the Canadian government, states in small

[3] Retrieved November 29, 2022.
[4] Retrieved February 7, 2023.

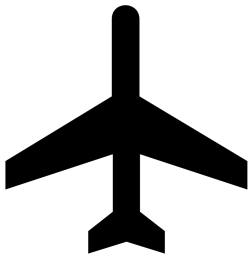
type at the end of a 12-page form that "your personal information may be shared with other organizations including the Canada Border Services Agency (CBSA), Citizenship and Immigration Canada (CIC), the Canadian Security Intelligence Service (CSIS) and law enforcement agencies," with no option to opt out of sharing your data, in presumed perpetuity, with Canada's domestic intelligence agency. Similarly, UCSIS's form I-485 ("Application to Register Permanent Residence")[5], familiar to green card applicants and refugees alike, notes on page 17 of a 20-page form that applicants must "authorize release of information... to other entities and persons where necessary for the administration and enforcement of U.S. immigration law." Immigration processes must obviously involve a level of due diligence to prevent safe harbor for criminals, but green card applicants and refugees are participating in the process from two entirely different levels of safety, and the casual breadth of consent specified on both these documents should be alarming to anyone.

Designers must architect workflows that not only minimize data intake but also provide clarity and nuance on data collection, particularly for people who are motivated to agree to any concessions while under duress.

As with the principle of avoiding predictive collection above, designers should never overreach by seeking consent for future unknown data disclosure opportunities.

Localize first, universalize later

Many designers strive for a universal syntax of usability, a commonality of form which engages users without reliance on language or geography. This is, broadly, a worthy goal. Iconography has long aspired to communicate without localization, and to step into any airport or rail hub in the world is to rediscover a common vocabulary of toilets, telephones and taxis based on the DOT pictograms developed for the US Department of Transportation in the late 1970s.



It's easy to forget, however, that universalization is often context-dependent: an icon of a baby in an airport might imply a nursery or changing table, but outside of an airport could be interpreted as a preschool. Cultural differences also become evident when context is missing: a cross may imply medical aid to some, but suggest Christianity or religious services to others. A thumbs-up has an implication of approval in North America (more so in the age of Facebook),

[5] Retrieved February 7, 2023.

but is simultaneously interpreted as an insult in Eastern Europe. A V sign made with the index and middle fingers can mean, alternately, peace, victory, the number two, or an insult analogous to the North American middle finger (which itself has alternate meanings in different countries). Color is also laden with cultural nuance: red denotes urgency in the United States where it can be found on stop signs and hospital signage, but in China has a positive association denoting good fortune, while in South Africa is the designated color for mourning.

For non-users, context will often be missing. Refugees and asylum seekers may arrive in a country that provides support services which were unavailable in the country they left. Designers should therefore aspire to universalize for clarity (while being wary of universalizing solely to save effort or cost), and must always remain grounded in, and aware of, localization. Data which is inexplicable or counterintuitive to non-users will remain of limited value to anyone.

Similarly, designers tasked with optimizing for analysis the data of non-users must remain alert to cultural nuances and the diversities of group identity, particularly when trying to accurately represent how non-users define their ethnicity, religion, gender or sexual orientation. Emigrants from over a third of the world leave countries where self-identifying as anything other than heterosexual risks a prison sentence or execution.

Assume third-party data is garbage

As discussed above, non-user data can be self-provided (first party), solicited (second party) or otherwise acquired (third party), and particular skepticism should be used with third-party data such as data purchased from data brokers, which is often inaccurate due to a lack of quality control mechanisms and an incentive to prioritize volume over precision. According to the FTC[6], at least 20% of consumer profiles maintained by third-party credit agencies contain material inaccuracies, and those inaccuracies tend toward a riskier interpretation of creditworthiness rather than a less risky interpretation. Even when data on non-users is accurate, it can be problematic; in 2020, the FTC sued consumer data provider Kochava for disambiguating and monetizing data from 125 million mobile devices to infer individual activity relating to churches, abortion clinics and domestic abuse shelters.

Designers should be aware of the provence of the data they make available for analysis, attributing it clearly and questioning its accuracy. Data flows which facilitate the migration of

[6] As reported to the US Congress in December 2012.

non-users into users should prioritize and facilitate the correction of erroneous third-party data. Disambiguating third-party data just makes a bad situation worse.

Final Thoughts

The science of UX is grounded in the art of ethnography: we watch users interact with objects and interfaces, and we analyze the footprints they leave in our analytics. The success of UX is defined by our understanding of user intentionality. This popular model of the willful user and the ever-observant designer only focuses on the visible peak of the iceberg, however. Our data-rich world is increasingly populated by individuals whose needs we can't infer because their data exists without their informed participation. This world of dark data, of unknown, unclaimed or inaccurate identity data, is an opportunity for designers to extend their reach into terra incognita, and the non-user experience is the lens through which designers can begin to look at people, not just at users or account holders.

Aside from being an opportunity for designers to broaden their craft, the non-user experience is also an opportunity for all of us to consider forward-looking mitigations for an ever-increasing volume of identity data. Personal data expansionism, to date, has been a challenge primarily in the purview of privacy advocates and policy-focused technologists, but it belongs equally in the worlds of design, urban planning, social science, humanitarianism and countless other disciplines where existing and new methodologies can be brought to bear in the service of those people we have an obligation to protect.