Digital Footprints

An Internet Society Reference Framework

JANUARY 2014

Table of Contents

Structure and Use of the Document

This framework document has been structured to make it simple for you to understand and manage your digital footprints. Each section deals with a particular topic: how digital footprints are created, why third parties are interested in our digital footprints, how privacy, economics, and legislation intersect, and so on.

Each section has been written to be self-contained, giving a description of the problem or topic, why it matters to you, and what you can do about it.

We have grouped the sections into three over-arching themes (economics, risk and context), so you can treat the whole thing as a single document, or pick out a single section that interests you most, or read the sections that related to a common theme.

We finish with a guidance section, giving examples of four types of action you can take to develop your understanding and control of your digital footprint. If that's your over-riding concern, just go straight to “How can I manage my digital footprints?”. However, the first recommendation you will find there is this: improve your understanding of the basic issues... which is exactly what we hope you will get from the rest of the document.

Guide to the Themes and Sections

Introduction: What is a digital footprint?

Theme 1 – Economics

Chapter 1. How did we start leaving such big footprints? (The role of cookies, and the effects of linkability)

Chapter 2. Is “monetized” the new “free”? (Advertising and the implicit economic bargain of “free” services)

Chapter 3. Who is tracking me, and how? (The commercial ecosystem of online tracking)

Theme 2 – Risk

Chapter 4. What problems can digital footprints cause? (The balance of social and economic benefit; linkability and contextual integrity)

Chapter 5. Different devices, different traces... (Apps, smartphones, and where we're headed)

Theme 3 – Context

Chapter 6. What dynamics are at work in the world of digital footprints? (Convenience, markets and the user choice shortfall)

Chapter 7. How does legislation affect digital footprints? (Issues of consent and cross-border data transfers)

Guidance: What can I do to manage my digital footprints?

What Is A Digital Footprint?

Digital footprints are the records and traces we leave behind us as we use the Internet. Your digital footprint may be a benefit or a risk to you... but the one thing it won't be is irrelevant. It's information that others use to make money, to find out what you like, where you go, and who you're having lunch with next Tuesday.

Your digital footprint can influence your online reputation and even your credit rating. It can mean you don't have to repeatedly log in or submit personal details to web sites. But digital footprints are visible to organizations with whom you may have no relationship, whose interests conflict with yours, and over whom you often have no control.

Most people are aware that when they share information about themselves on the Internet, such as with social networking services, and when they use on-line services, such as electronic mail, instant messaging, or voice calling, they have given up some control over their privacy. As recent controversy in the US and elsewhere has highlighted, the information we entrust to others—even when we think it is private—is out of our control.

This loss of control is frequently the result of explicit acts: making a Skype call, sharing something on Facebook, uploading pictures to Tumblr, sending an email to a Hotmail user. We may expect *some* privacy, but we know we've given something up, and we've left a clear imprint at each of these individual services.

But what about the trail we leave implicitly, as we travel around the Internet? Is it possible for someone to follow us around in the virtual world of the Internet, tracking our digital footprints, tracing the impressions we leave? The answer is "yes." Your digital footprints are bigger than you may have thought, and they are being used—usually for commercial purposes, but sometimes for other reasons—to track you, customize for you, and market to you. In short, your digital footprint is a monetizable asset... but seldom to you.

Theme 1 - Economics

Chapter 1. How Did We Start Leaving Such Big Footprints?

"Mr. Holmes, they were the footprints of a gigantic hound!"

~ A. C. Doyle, *The Hound of the Baskervilles*

Digital footprints should be a significant privacy concern for Internet users, because they can be used to track user actions and are a basis for "profiling" by online service providers and others. Over time, the

technology to create profiles of Internet users has become increasingly sophisticated. Few users realize how extensive their digital footprints are, and or how commonly the resulting data is shared by third parties.[1]

The explicit footprints we leave as we participate in Internet conversations can be obvious to us, if we pay attention to them. For instance, if you Tweet that you have just arrived in Sydney and the sunset is spectacular, those are quite explicit disclosures about where you are and when (assuming you're telling the truth). But what about the implicit footprints? Each time you visit a web site, you reveal some information about yourself to the owner of the web site: your IP address, which may include your geographic location, your web browser type and operating system, and, often, the last web site you visited. These bits of information seem relatively innocuous and even fairly anonymous. If these are footprints, they are – individually - fairly light ones.

In fact, these footprints can pose a problem to some people because they're too light. Internet services such as online commerce, social networking, and web mail require that a web site be able to link up multiple interactions, such as putting a book in a shopping cart, and later clicking "Pay Now". They need to know when the person doing something now is the same person who did something previously – and often this is in the user's direct interest, too. IP addresses won't do the trick, because several people might be using the same IP address at the same time: something more is required. One solution to this problem is the cookie. A cookie is one way of tying multiple actions by a single user into one connected stream; as it becomes possible to link more actions together, so the potential privacy impact on the individual starts to grow, even if the individual data points are not especially revealing.

Digital footprints in the form of cookies are used to make the Internet more usable, and can also help make individual transactions more secure. A great many current Internet services are designed to rely on the availability of cookies, and cannot function – or function fully as intended – if cookies are blocked. In other cases, cookies are used entirely for the convenience of the web site and bring no real benefit to you.

The positive side of digital footprints is an important factor in our consideration of privacy. However, cookies are not the only mechanism to provide security and persistence. Transaction security decisions may rely on a combination of factors, including cookies but also other mechanisms such as "decorated URLs", browser ID strings (referred to as the "user agent" string), the user's IP address, and so on. Web developers have settled on cookies as one of the most convenient ways to add persistence and security to your web experience, which is why they are ubiquitous. Let's look at cookies in a bit more detail.

A cookie is an arbitrary string of letters and digits - something without any inherent meaning - that a web site sends to your web browser. Here's an example of a cookie:

JSESSIONID=0000e7B1glDiG4yqQy4Rivr5rCf:17q9uijvp

Your web browser stores the cookie when requested, and then every time you revisit the web site, the web browser sends the cookie back to the web server. Although the cookie usually doesn't have any special

1 The Wall Street Journal has published a "What They Know" series of documents exploring this topic in greater depth. Interested readers can start at http://online.wsj.com/public/page/what-they-know-2010.html to learn more from this series.

meaning to you, the web site can store profile and preference information in the cookie, use the cookie to point to stored profile and preference information elsewhere, or use it to record things like when you last authenticated.

Browser cookies work behind the scenes to provide continuity and persistence. For example, without cookies, you might have to type your username and password over and over as you read your webmail, browsed an e-commerce site, or participated in a social network. Not just once per site, but again and again for every single page.

Web sites generally set a cookie in your browser at the instant when you first visit the site. The browser stores the cookie behind the scenes, and then sends the cookie back to the web site every time you click. One result is that the web site can stitch together your actions to improve your user experience—even if they're separated by days, weeks, or months. Most users don't think about it, but by default, their web browsers contain thousands of cookies, placed there by each web site they visit. However, this also means you're leaving bigger and bigger footprints. Cookies don't just link up transactions; they enable web sites to keep track of you every time you visit.

Most cookie-setting service providers explain this in terms of 'optimizing their relationship with the customer'. But if the service provider holds all the data – often without the user knowing or seeing what is held – the resulting relationship is rather like watching someone through a two-way mirror. And that can be disconcerting.

> "Every step you take,
> Every click you make,
> I'll be watching you...
> Oh, can't you see? You belong to me..."
>
> *~ With apologies to Sting and the Police*

As each one of us uses the Internet, our wanderings through web sites, search engines, social networks, and electronic mail leave information about our professional and personal tasks, commercial activities, and how much we like cats and Justin Bieber videos. If a service provider holds account information for you, such as your email address, payment details, purchase history or other personal information, the cookie links everything you do with this information. The concept of linkability is a key one in any analysis of online privacy, because linkability does more than almost anything else to erode users' ability to keep personal data within a single context, and thus to manage their own privacy.

Each individual footprint is small, but when linked they can form a surprisingly complete profile about us. When web sites decide to share this information with each other, it becomes possible to build a profile of you, using raw data such as the web sites you've visited, the products you've bought or searched for, your address, and any other bit of information you've given to any cooperating web site: age, sex, health, marital status, employment, financial information … the list is as long as everything you've ever shared on the Internet. In fact, it is longer – because based on the raw data, profiling companies make inferences about your habits, preferences, values, aspirations, and even your intentions and future behaviour.

We discuss ways to minimize digital footprints later in this framework document for Internet users who want to be in greater control.

Chapter 2. Is "Monetized" the New "Free"?

"There is one thing
I mean everything has a price
I really hate to repeat myself
But nothing's free"

~ Alice Cooper, "Gimme"

The heavy use of digital footprints to track users and customize content is an outgrowth of the basic economic bargain of the Internet. Because most of the infrastructure of the Internet is funded by marketing in some form, publishers and marketers exploit digital footprints to target their products at the most appropriate audience.

Digital footprints would exist even if there was no commercial need for them, but the commercial side of the Internet has capitalized on the opportunity they represent. Advertisers and marketers have grown dependent on the power digital footprints have given them to observe, link and mine data about Internet users.

Online services are not truly free of charge, and never have been. A lot of the content and services seem to be free, in the sense that we don't directly pay for them. There are exceptions—some newspapers and magazines, pay-per-view video streams, and chargeable information services such as industry analyst reports —but for the most part, there is no apparent cost to view data on a web site, read someone's blog, watch a video, post a picture, or join a social network.

But the word "apparent" is significant: even if we're not paying directly, we are paying indirectly. Someone has to fund the servers, the data centers, and the networks that underpin online service provision. Originally subsidized through government research grants, the Internet is now subsidized through a powerful economic force: marketing. Of the top 100 web sites (by traffic), only one—wikipedia.org—is completely free of advertising.

The phrase used to describe this is: "If you're not paying for the product, you are the product." It's not a new idea, and it's not a black-and-white distinction, but it sums up most "free" online resources very succinctly. If you don't pay a subscription fee for a service or application, that service is funded by monetizing information about you, your social circle, and your collective interests and preferences. Sadly, the converse is not also true: the fact that you pay a subscription for a service, or a premium for an "ad-free" version, does not guarantee that your personal data is not collected and monetized.

For almost every Internet site, every time you look at a web page, someone has an interest in showing you an advertisement.[2] The cost of delivering an advertisement on the Internet is very low compared to other mechanisms, such as billboards, newspapers and magazines. This means that consumers and the advertisers who want to reach them pay for most of the "free" content on the Internet— and leave a healthy profit margin... for someone. Google, for example, had a profit of almost USD $11 billion in 2012, almost all of it from placing advertisements.

This trade of your eyeballs for their servers, networks, and content is the essential economic bargain underlying most of the Internet. This is not a secret – and yet few people who use the Internet focus on the bargain, rather than the convenience of the services they want. Even fewer actually modify their online behaviour *because* the bargain causes them concern.

But from the perspective of the marketers, the Internet offers both opportunities and challenges. The opportunity is of cheap, direct access to receptive consumers. The challenge is that the Internet is so information-rich that it can be hard sorting the receptive consumers from the rest. In the old days, advertisers had a pretty good idea of who was reading "Modern Bride" magazine and what they might be buying in the next few months. But maximizing return from the Internet requires a more holistic view of the consumer than that, and in a much more diverse global market. When Facebook sells a spot on a web page to an advertiser, that spot can be tremendously valuable or practically worthless - all depending on who is looking at the page, what their interests are, and whether they're in the mood to buy what's on offer. This gives advertisers and the publishers who place their ads a strong incentive to find out as much as they can about their audience. This helps them identify the right demographic, the right language, the right product, the right time, and all the other factors that

As mentioned above - even if you are paying for a service, that doesn't mean you're not being tracked. Some web services do it for everyone, just out of habit or because it's simpler and cheaper to track everyone rather than work out how to do it selectively. And others are trying to make even more money, just as a magazine or television channel you pay to receive still has advertising. They may not be showing you advertisements, but they may still be selling information about what you are doing to someone else---and that third party may want to show you an advertisement.

These are powerful forces: advertisers' desire to customize ads to the audience; publishers' wish to charge the highest possible sum for showing someone an advertisement, and the incentive to track the buyer and maximize commercial return. They create an overwhelming incentive to collect mine, re-sell and monetize data consumers, and power a commercial engine over which the individual consumer has little or no influence.

2 If they aren't showing you an advertisement at this moment, they may be gathering information about you and your interests for someone else…who wants to show you an advertisement or sell you a product or service.

Chapter 3. Who Is Tracking Me, and How?

"Data is everywhere. It exists.
We're just pulling it into one place…"

~ Fahad Hassan of "Always Prepped," November, 2012

To understand digital footprints better, it's helpful to know the key players that are collecting footprint data and following us around the Internet. In the scope of this paper, we're just going to look at the commercial use-case.

In the world of online advertising and tracking, three main players work together to track users and create composite profiles: advertisers, aggregators,[3] and publishers.

By "publishers", we mean the companies that publish advertisements online, pairing advertisements with web page content, games and so on. Any on-line magazine falls into this category, as do search engines, blog publishing platforms, and millions of other sites that provide a vehicle for publishing advertisements alongside content.

By "advertisers", we mean the companies that market consumer products and services. The people who have something they want to sell you.

In many cases, advertisers work directly with publishers. For example, a car manufacturer might ask an online newspaper to publish ads for their latest model on the news site for viewers in Italy, knowing that this is a strong target market for them. Then there's the part of the ecosystem that you may never have seen.

Data aggregators and brokers, companies with names you've never heard of like BlueKai, Gravity, Rio, OutBrain, and Dataium, get involved when the advertiser wants to be more specific. Data aggregators collect (supposedly) anonymous data from their partners and use it to target ads.

Let's walk through two examples showing how cookies are used to track you. Google's News (news.google.com) is a news aggregation site, providing pointers to the top news stories from thousands of news sources. If you have no relationship with Google and you simply click open the web page, you'll see the top stories, customized for your current location (such as "Tucson, Arizona") based on your IP address. At that moment, Google News will also place a tracking cookie in your web browser; the browser will store that cookie on your device. The browser's job is to send that cookie back to the web site each and every time you return to it or ask for another page.

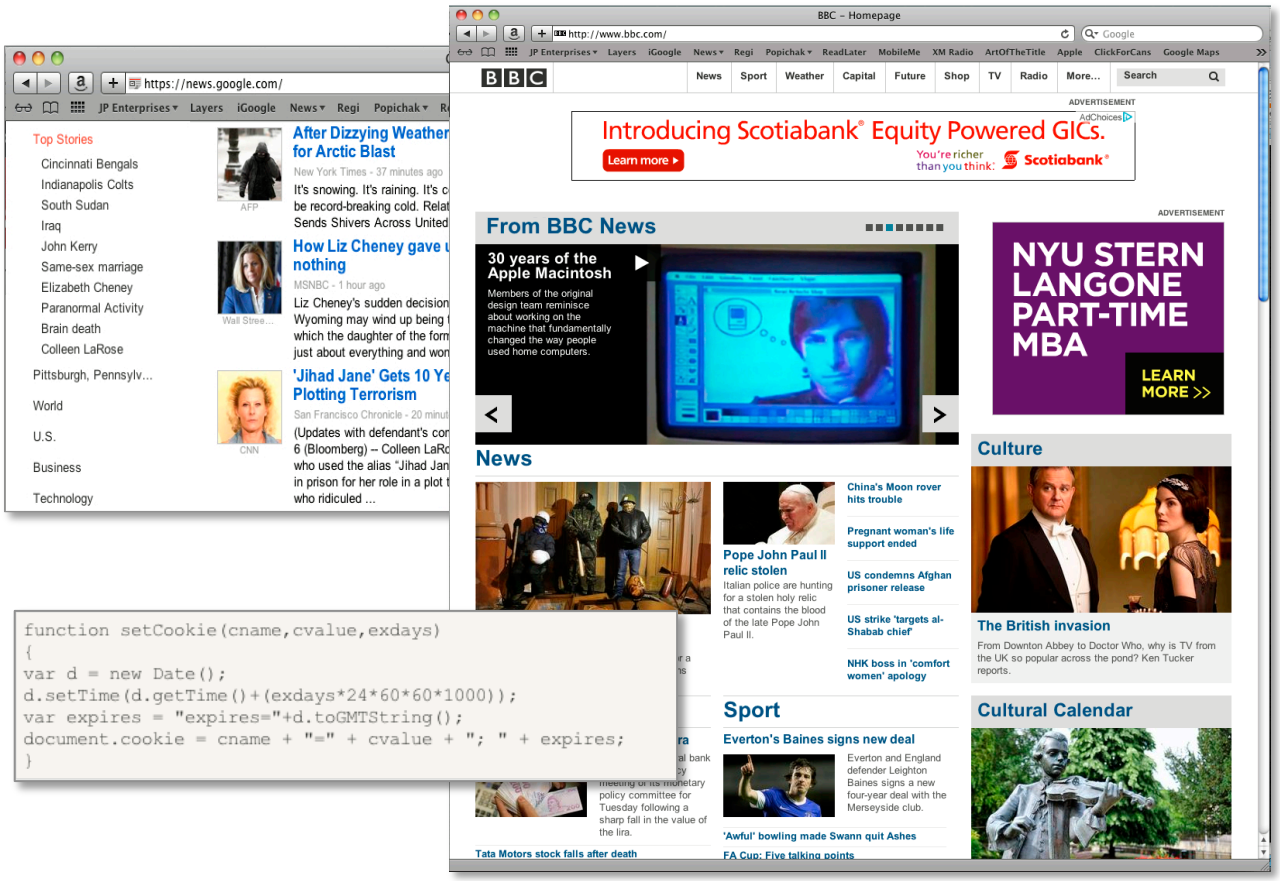
If you click on a story on the Google News page, and that story is hosted on another web site, your browser opens up a new window to display it. Then, your browser connects back to Google News, sending it the

3 Aggregators are often also called data brokers.

tracking cookie as well as the URL for the news story you requested. Thus, Google News learns about every story you click on, all linked up by the original Google News tracking cookie.

Cookies are attached to domain names. When Google News sends down a cookie to your browser, it uses the wildcard domain name "*.google.com" rather than the more specific "news.google.com." This means that the tracking cookie is sent back to any web site that ends in "google.com," not just Google News, but all Google services.

When the story you requested pops up, the news service web site knows that the request came from Google News (because your browser sends this information, the "Referer," as well). Although the news service web site doesn't get the tracking cookie from Google News, it will generally send its own tracking cookies to your browser. And if any of *those* web pages have advertisements from third parties on them, the third parties will also send their own tracking cookies.



For example, (as of when this paper was being researched) if you clicked on a news story from the BBC News site, your browser would pick up a dozen or more cookies from the BBC, a couple from atdmt.com, and several each from doubleclick.net, mediaplex.com, and revsci.net.

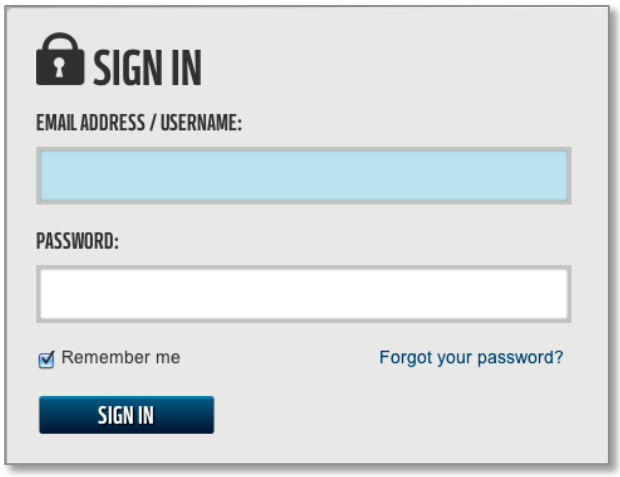
In theory, each of web sites operates independently and your browser will only send back cookies to the web site that set them. Thus, the cookies from Google News are not available to the BBC, and from the BBC to doubleclick.net. However, the different web sites can communicate in other ways, both through your web

browser by leaving clues for each other (such as in the URL or cookies set for other domains) or simply through a separate process of information sharing.

Up to this point, while the web sites you're using have a lot of information about your interests, or at least what news articles you read, they don't really have any information about who you are or your profile. But even limited information may allow someone to guess your age, sex, income and other details.

So far, we have just looked at the example where you read news stories without authenticating to any of the web sites in question. Now let's add some more depth to the example. This time, we'll assume that you have started by logging in to an account you created on one of your favourite sites.

For instance, let's suppose you've created an account on a motor-sports enthusiast web site, and identified yourself as 32 years old, female, and living in Colorado. The motor-sports web site places a cookie in your web browser, as before, but this time it also leaves a "third party cookie" with data for the aggregator. This is a cookie left by one web site, but containing the URL (the "return address") of another – in this case, the aggregator. In this case, the motor sports site contributes what it knows about you from your profile: "cookie #117555 identifies a female, 32-year-old, from Colorado, who is interested in motor sports". As in the previous example, cookies are just one way in which this information might be exchanged between the two parties.



Any other site that embeds content from the same aggregator could use the same mechanism: your later visit to a financial services site might generate additional information, letting the aggregator know that cookie #117555 searched for information about loan rates, and a third web site might add report that cookie #117555 looked for the best price on baby food.

The aggregator puts all this together and profiles the user with dozens of different criteria: age, income, shopping habits, sex, location, interests, you name it. Aggregators work with publishers to sort viewers into different categories, and then offer advertisers the prospect of access to those "eyeballs". Then, when the maker of a new baby car seat for sports cars wants to reach customers, cookie #117555, gets shown their advertisement the next time she goes to a web site allied with that aggregator.

For a few very large publishers, the jobs of aggregation and publication are handled by the same company. Aggregators have a lot of data on a lot of people: in early 2013, BlueKai had information on 85 million unique users, while OutBrain offers up 364 million, and Rio claims nearly 500 million.

The assumption that third-party tracking cookies are anonymous is widely disputed. Although most aggregators take pains to not store information that directly identifies a person, researchers have shown that this is not much protection. It only takes a few data points, especially when location is included, to track a

set of web events back to a real person—which offers the possibility of following that real person both back and forward in time.

For the most part, digital footprints are used in a commercial context by companies who want to market products and services to us. In that context, the visible result is generally an innocuous, but slightly creepy, impression that someone is following you around and trying to sell you stuff tailored to your apparent interests. But digital footprints can also result in a loss of anonymity, through the sharing of information by third parties who have little regard for the consumer's privacy.

Theme 2 - Risk

Chapter 4. What Problems Can Digital Footprints Cause?

"On the Internet, no one knows you're a dog."[4]

One of the Internet Society's core tenets is about maximising the economic *and social* value of the Internet. What you may have noticed in previous sections of this framework document is a focus on digital footprints as part of the *commercial* Internet ecosystem. One of the side effects of digital footprints is a loss of privacy and anonymity online: from the Internet Society's broader perspective, this undermines the social value of the Internet.

As we participate in various Internet activities, such as sharing in social networks, reading and sending electronic mail and instant messages, and making calls using Internet telephony, we are leaving behind evidence of what we've done, where we've been, what we've been thinking, who our friends and families are, and more. These footprints build up over time, and can become enormous.

The implicit footprints we leave behind also can be used to track us, and to link information we have explicitly shared, in one context, into a larger and more complete profile that extends across the contextual boundaries of what we do online. The privacy implications of linkability are profound. It is one thing to discuss details of an ailment with your doctor (with all the contextual rules that implies) and quite another to see the same information published on a "comic ailment of the month" blog. The ability to keep different contexts separate when we want to is a vital part of personal privacy, on- or offline.

In an era of "big data" analytics, organizations—not just governments—are able to analyze huge amounts of data from our footprints and link it across multiple contexts.

When an advertisement pops up on a web page for an object researched two days ago on a different site, it's a sign that someone has been sharing our activities with the advertiser. If we expected those two contexts to be separate, we are likely to feel that our privacy has been violated. These specific examples about contextual integrity are illustrations of a broader problem.

---

4 The cartoon from which this caption comes was published 20 years ago: 30th July, 1993.

The Universal Declaration of Human Rights offers everyone a "right to privacy," but there is no universal agreement how privacy works on the Internet.[5] Privacy and other human rights are social conventions – and subtle and complex ones, at that – and the technology of online service delivery is still an immature and unwieldy way to express those conventions in our digital lives.

Although the Internet and its associated commercial services happen to have developed in a particular way, we should not assume that that is the only way, or that it represents the healthiest possible balance between commercial and social benefits. There is scope – and, some would say, pressing need – for improvement.

For instance, in addition to privacy, there are other areas in which individuals' interests are potentially put at risk because of their digital footprints. One such area is anonymity. There are circumstances under which people feel free to express themselves openly, only to the extent that they can do so anonymously or pseudonymously. This may be because they wish to say things that might be too dangerous or sensitive to express identifiably. As things stand, we have little option but to trust third parties to respect our preferences regarding privacy (even if, most of the time, we have no tools with which to express those preferences).

It takes very little information to tear the veil of anonymity. The linkages between digital footprints, IP addresses, phone numbers, e-commerce, and on-line activities all make it possible to ascribe "anonymous" actions to a real-world identity. These links can often be made by parties whose interests run counter to those of the individual trying to remain anonymous.

When footprints are shared between on-line merchants for advertising purposes, the violation of privacy and reduction in anonymity is usually little more than an annoyance. However, if those footprints are linked together and matched to an identity by someone more official or more malicious, then there is a real threat that on-line activities can have significant detrimental consequences for the individual in question.

Penetrating on-line anonymity can strike in both directions. For example, in February of 2011, Aaron Barr, CEO of HBGary Federal, told the Financial Times that he used social media footprints to identify members of the "Anonymous" hacking group, and would publish their real names. Before he could do this, someone claiming to represent Anonymous broke into his employer's servers and published thousands of email messages and other documents, resulting in significant economic and reputational loss for parent company HBGary.

The loss of privacy and anonymity reduces public trust in the Internet and harms the entire Internet community.

5 Or even an agreed definition of what "privacy" means.

Chapter 5. Different Devices, Different Traces...

> "The smartphone revolution is under-hyped. More people have access to phones than access to running water. We've never had anything like this before since the beginning of the planet."
>
> *~ Marc Andreessen, 1/May/2012 interview at Wired Business (http://en.wikipedia.org/wiki/Marc_Andreessen)*

Laptops and desktop computers tend to leave a very different footprint from smartphones and tablets. Modern smartphones (and tablets, which are closely related) have evolved as creatures of the modern Internet ecosystem, and as a result often work in ways that create a more intrusive footprint.

The standard web browser is very different from the applications ("apps") that smartphones and tablets use. Apps connect directly to Internet services (and, indeed, to other apps and other devices) using specific interfaces (by contrast with the much more generic interfaces that are used by browsers). Control over which information is sent to other services/devices rests in the hands of the app developer, and is exposed to the end user only to the extent the developer permits. Mobile devices, in particular, also give users less ability to connect anonymously.

Because smartphones generally are location-aware, it is easy for services to tag your activities to your location; location services are often either enabled by default, or requested by apps when you first install them. The location data can then be shared explicitly, if the application retrieves your location data and send it to the Internet service, or implicitly – for instance if the pictures and videos you upload were tagged with the location, date and time they were taken.

Smartphones are also designed as very personal devices, not shared between friends and family members. With unique serial numbers (such as the IMEI[6]) inside of each phone, smartphones can link the real identity you have provided to your mobile phone carrier with all of your Internet activities, bridging the physical and virtual worlds. Because many countries now require some type of identification to register all mobile phone subscribers, it becomes simple for third parties, such as law enforcement agencies, to link Internet activity to a particular smartphone and thus to an individual user.

6 **International Mobile Station Equipment Identity (IMEI):** Defined as an International standard, the IMEI is a unique number programmed into each individual phone by the manufacturer. The IMEI is supposed to be permanent and unchangeable by the end user. (3rd Generation Partnership Project technical standard TS 22.016, http://www.3gpp.org/

Of course, mobile devices are not unique in identifying the subscriber. The same principle of some type of authentication and registration applies to payment, banking, and Internet services, making linkage between on-line and real-world activities easy for anyone who has access to the registration data. What may then distinguish the two cases is the extent to which a given country's laws control service providers' behaviour. This may be in the form of privacy regulations that apply to a specific industry (e.g. financial services), or to the use of personally identifiable data regardless of which industry is involved. Or both kinds of regulation may apply; for instance, in the United Kingdom there are general laws about the processing of personal data, and then additional regulations specific to financial services.

Smartphone vendors, aware of the potential for abuse of these features, generally at least implement controls over whether location data are shared, and to block the use of device-specific identifiers by applications. Some controls over sensitive information, however, are based on based on settings at the device level, and others at the application level. These capabilities vary significantly by smartphone platform; so does the level to which they are documented, and the ease with which the average user can discover and manage them.

But once a user starts taking tagged pictures, or gives permission to a newly installed application to see location information, the permission granted to the application is rarely revisited. An application may ask once “may I use your location,” but doesn’t remind you every few weeks that you gave it permission. And not every application is well behaved, or every operating system bug-free. Even a diligent consumer who regularly checks their privacy preferences is sharing more information than they have permitted. For example, simply turning on a mobile phone allows the carrier to locate the phone to a certain degree of accuracy. And if someone else “tags” you in a photograph they’ve uploaded to a social network, your location at a particular date and time have just been shared on the Internet by your friends—even if your own smartphone is switched off and lying on the kitchen table at home.

In some cases, it is the device vendor who may have access to sensitive and private information on the smartphone. For example, it was discovered by a researcher in June, 2013 that Motorola’s “Blur” service platform has stored profile, credential, traffic, and device status information for recent Motorola phones running the Android operating system - with automatic updates by the smartphone every time the user makes a configuration change.[7]

In practice, then, an individual's privacy can be affected by the actions of many other entities. Here's a list just from the scenario of mobile device use:

- Device/smartphone vendor
- App developer
- Network operator

7 Lincoln, Ben, “Motorola is Listening,” retrieved July 4, 2013, http://www.beneaththewaves.net/Projects/Motorola_Is_Listening.html

- Operating system developer
- Internet Service Provider (ISP)
- Online service provider (e.g. retailer, social network)
- Friend/acquaintance... or their app
- Other device

Users of desktop computers and laptops primarily leave footprints via their web browser. The comparatively mature controls offered by browsers themselves, or by supplementary plug-ins, make it easier for the end user to control what is shared, and to clear out identifying information such as cookies that might otherwise reduce personal privacy. Desktop computers also leave "muddier" footprints, as users often run multiple web browsers, and the devices themselves are somewhat more likely to be shared.

If desktop computers have a privacy edge over smartphones, it is possible that they will not keep it for long. For instance, as user expectations are driven more and more by the smartphone/tablet experience, the pressure is on for operating systems like Windows 8 to act more like a smartphone than a 'traditional' desktop computer - with the same potential loss of user control and privacy.

Internet users concerned about the digital footprints left by their smartphones need to take an active role in managing their privacy settings. The task of carefully monitoring and controlling privacy poses a significant overhead and may be more complex than many smartphone users expect. The challenge for all of us, as consumers and users, is to recognize the value of our personal information and our privacy: only by adjusting our values and, as a result, our behaviour, can we hope to make better, sustainable decisions about privacy.

Theme 3 - Context

Chapter 6. What Dynamics are at Work in the World of Digital Footprints?

> "I hope we will use the Net to cross barriers and connect cultures"
>
> ~ *Tim Berners-Lee, (*http://en.wikipedia.org/wiki/Tim_Berners-Lee)

The online world reaches into our lives so pervasively that it is often hard to untangle the dynamics that govern its behaviour. However, three themes stand out from the rest: the cultural dynamic, the economic dynamic, and the dynamic of convenience.

Online services, while accessible worldwide, must all originate somewhere, and that often infuses them with a particular cultural perspective. Each country brings to the Internet brings its own cultural norms, legislative

and regulatory models, and economic frameworks. As a global resource, the Internet has become a fascinating study in contrasts.

Digital footprints, which touch on issues of personal privacy, data sharing, end-user control, and anonymity, produces different reactions across the diverse constituencies that make up the Internet. What is acceptable and common to one group is sometimes unacceptable and unusual to another. This is part of the nature of the Internet.

To some, the answer seems simple: if you don't like the privacy model of a particular Internet service, choose an alternative. End users can vote with their clicks by avoiding services that don't meet their expectations.

However, this advice only works if two things are true:

- Internet users are aware of the privacy and data protection implications for every service they use; and
- a true choice of services is available.

Our collective experience is that neither of these statements is always true. Not every Internet user knows how services are sharing their information, not every service has a substitute, and sometimes all the alternatives suffer from the same drawback.

Even Internet users who are very active in controlling the digital footprints they leave have no option but to rely on imperfect knowledge. Sometimes that is a direct result of choices made by the service provider. For instance, it is in the interests of a social network service to encourage users to ignore the fact that everything they do within their social circle is being inspected and monetized by a third party. The economic dynamic gives the service provider a strong incentive to collect data, and to keep users under-informed about that aspect of the service.

And while the Internet is incredibly diverse, there is not always a choice of services available. In some areas, such as social networks, even when competing services are available, other factors might make the competitive service unattractive – a social network will have little appeal if none of your friends is using it.

Finally, there is the dynamic of convenience. Most of us would rather use something which is convenient but privacy-eroding, than use a product that makes life less convenient for us. Our preference for the "convenient" option is also strengthened if we see no evidence that our privacy is being eroded. Like many kinds of human behaviour (smoking, eating fatty food, poor posture), if we cannot see an immediate harm from our actions, we tend to assume they aren't harming us. The combination of convenience and lack of apparent harm lulls us into privacy-eroding habits. As with any other habit, our chances of changing our behaviour depend on the value we place on our privacy, relative to the "convenient" alternatives.

There is no simple answer. The first step is to acknowledge the different cultural models at work, and understand that Internet users come to the table with different backgrounds, expectations and values. Open discussion between significant stakeholders can help to educate both end-users and service providers about

these concerns. Ultimately, though, it is up to us as individuals to be clear about the choices we make, the values those choices represent, and the outcomes we face as a result.

Chapter 7. How Does Legislation Affect Digital Footprints?

> "Good people do not need laws to tell them to act responsibly, while bad people will find a way around the laws."
>
> *~ Plato*

The Internet is global, but privacy laws are not.

Most privacy laws and guidelines focus on "personal data" or "personal information", often defined as "information relating to an identified or identifiable individual." However, definitions vary and continue to evolve. For example, there is growing awareness of the potential privacy impact of *any* information that can be used to single out or treat an individual differently, even if the individual cannot be identified.

As a result, there are proposals to make this explicit, either in the language of privacy regulations or in explanatory materials[8]. Combined with advances in data linking, storage, retrieval, correlation and analysis, ever increasing amounts and categories of data are likely to fall within the scope of privacy and data protection laws.

Generally speaking, privacy and data protection laws only apply to information about living individuals, but some countries extend the application of the law to information about deceased individuals[9]. The more "digital" and online our lives become, the more significant the question of managing an individual's "digital legacy" after death becomes.

Some countries have laws that are designed to protect pre-defined classes of data that are traditionally considered more sensitive, including medical data, financial data, and government issued identifiers. However, these laws were developed when the boundaries between data classes seemed clearer. For example, should the results of a Web search on "flu" be considered medical data? What about "do I have HIV" or "am I pregnant?"

Article 12 of the Universal Declaration of Human Rights, and Article 8 of the European Convention for the Protection of Human Rights and Fundamental Freedoms both refer to the individual's right to respect for their

8 For example, the Article 29 WP Opinion 8/2012 proposes the following definition for the new EU Data Protection framework: "any information relating to '…an identified natural person or a natural person who can be identified, directly or indirectly, **or singled out and treated differently**, by means reasonably likely to be used by the controller or by any other natural or legal person, in particular by reference to an identification number, location data, online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that person'…"

9 For example, parts of the Singapore Personal Data Protection Act 2012 apply to individuals who have been deceased for less than 10 years (see http://statutes.agc.gov.sg/aol/search/display/view.w3p;page=0;query=CompId%3A32762ba6-f438-412e-b86d-5c12bd1d4f8a;rec=0;whole=yes)

privacy or private life, but there is no privacy or data protection law that applies everywhere in the world: no single set of data processing rules that covers all Internet services and users.

There is fairly widespread international agreement on a set of key principles[10] (i.e. collection limitation, purpose specification, use limitation, etc.), but the practicalities of privacy law and enforcement vary widely by country. Some countries and regions, such as Europe, take a rights-based approach towards data protection and privacy. Others, even if they do not take a rights-based approach, have adopted a "comprehensive" approach to privacy. Yet, others, such as the United States, rely on more industry-specific laws, self-regulated best practice and codes of conduct. Then, there are the countries that have no (or only rudimentary) privacy laws.

These differences all increase the challenge of bridging the gap between country-specific laws and the frontierless nature of the Internet.

To enable cross-border flows while protecting privacy, a number of groups of countries have reached binding or non-binding agreements, such as:

- OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data[11], specifically Part 3
- Council of Europe Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data[12], specifically Chapter 3
- APEC Privacy Framework[13] and APEC Cross Border Privacy Rules system[14] (a voluntary accountability-based system)
- U.S.-EU & U.S.-Swiss Safe Harbor Frameworks[15]
- EU Binding Corporate Rules[16] (for multinational companies).

Some privacy and data protection laws, such as the EU Directive 2002/58 on Privacy and Electronic Communications, specifically target data associated with Internet use. Some of the data addressed may not fall within the legal definition of "personal data", but nonetheless has a privacy effect – for instance if our browsing habits are being monitored.

10 http://oecd.org/internet/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm

11 http://oecd.org/internet/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm

12 http://conventions.coe.int/Treaty/en/Treaties/Html/108.htm

13 http://www.apec.org/Groups/Committee-on-Trade-and-Investment/~/media/Files/Groups/ECSG/05_ecsg_privacyframewk.ashx

14 http://www.cbprs.org

15 http://export.gov/safeharbor/

16 http://ec.europa.eu/justice/data-protection/document/international-transfers/binding-corporate-rules/index_en.htm

However, if the laws are too technology-specific, they may have the unintended consequence of shifting the unwanted activity outside the scope of the law. For example, Article 5(3) of the EU Directive mentioned above attempted to regulate HTTP cookies and similar mechanisms (e.g. flash cookies, DOM storage) that were identified as a privacy-threat when the Directive was adopted. This kind of approach may, perversely, give service providers an incentive to look for other, unregulated means to monitor or profile users, such as browser fingerprinting and server-based storage. These may prove harder (or impossible) for the user to detect. Adoption of the EU's "cookie-regulating" measures remains low and, arguably, ineffective. Only a small number of countries, including UK, France, Belgium, Poland, Italy, Spain, Sweden, and the Netherlands, are actively requiring websites to obtain user consent to use tracking cookies, and even in those countries, few sites have implemented the regulation in a way which genuinely gives users the hoped-for level of control.

Also, there is a case to be made for uses that that genuinely improve the users' browsing experience and/or security (for instance, if cookies are used in support of a two-factor authentication protocol).

A critical regulatory factor is the issue of consent. User consent plays an important role in extending collection, use and disclosure of "personal data" beyond what is strictly necessary to provide a product or a service. One approach is to insist that decision-making must rest with the individual most likely to be affected, but putting this into practice this can be more problematic than one might expect. Internet users:

- rarely have the information or understanding they need to make an informed decision;
- are often given choices that are binary (say "yes" or you don't get service);
- may have only uncertain knowledge and incomplete information on the potential consequences of consenting; and
- are increasingly being asked to disclose the personal data of other people, such as when a service asks for your contact list.

Most Internet browser software offers users the option of sending a message to websites they visit that they do not wish to be tracked. However, so far, relatively few websites have said that they will honor users' requests.

In summary: the Internet ecosystem is global and complex, and regulating it is a challenge. There is no single set of rules, and no single definition of what data needs to be protected. Regulating at the technology level is generally unsuccessful – but regulating for behaviour makes the law highly culturally-dependent, and harder to reconcile with other jurisdictions. And the issue of user consent seems simple on the surface, but conceals deep technical and even behavioural complexity.

Realistically, we cannot hope for a one-time legislative fix to the privacy problem: we should expect to have to engage in a continuous process of evaluation and adjustment.

How Can I Manage my Digital Footprints?

> "You have zero privacy anyway. Get over it."
>
> *~ Scott McNealy, 1999* [17]

Managing your digital footprint on the Internet takes thought, time and effort. It involves struggling against our own inertia in the face of convenient but privacy-eroding defaults, and against the concerted, persistent efforts of organisations that have a financial interest in persuading us to sacrifice our privacy in the interest of their profit. You probably only have limited time and energy to devote to what seems like an incidental task, while the advertisers and publishers—well, that's the only thing they have to do all day long, and they're pretty good at it.

Rather than try to fit detailed practical guidance into this framework document, we first describe four levels at which you may choose to take further action, and then give four more specific examples. We also include several pointers to other sources of more detailed advice.

1. Improve your understanding of the basic issues

Reading this framework document is a good first step. Think about the implications of everything shared on the Internet being a privacy risk to some degree. Review the ISOC tutorials on Digital Identity and Privacy,[18] and other sources of instructional material.[19]

2. Develop your 'basic hygiene" habits

Privacy is a contextual thing. If you use different "personas" for different aspects of your online life – whether that's one email address for work and another for home, or a one credit card for online shopping and another for everything else – it will help keep different parts of your digital footprint separate. Be mindful about what you share via social sites and elsewhere, because that data is probably more public and persistent than you might anticipate.

---

17 White male Silicon-valley executives have such a wonderful track record of outrageous quotations that it is difficult to stop with Scott McNealy's, co-founder of Sun Microsystems thoughts, as one should also include Eric Schmidt from Google ("If you have something that you don't want anyone to know, maybe you shouldn't be doing it in the first place") and Mark Zuckerberg of Facebook ("*People have really gotten comfortable not only sharing more information and different kinds, but more openly and with more people. That social norm […] has evolved over time [...] we decided that these would be the social norms now and we just went for it").*

18 "Manage Your Identity," a series of three five-minute tutorials on online identity, identity protection, and privacy protection. http://www.internetsociety.org/manage-your-identity

19 The Jericho Forum prepared a series of 4-minute videos on Identity (http://www.youtube.com/watch?v=6FHGe8yHeQE), Operating with Personas (http://www.youtube.com/watch?v=eK4-dh8I_Dk), Trust and Privacy (http://www.youtube.com/watch?v=1FFxCkWh9Ho). The bigger picture of Identity (http://www.youtube.com/watch?v=CI_9mIZSshg), and Building a global identity eco-system (http://www.youtube.com/watch?v=3aDlAAPd4v0) which provide a good background in the area.

3. Become a sophisticated user of your online tools and services

Very often, the default settings for browsers, devices and apps are set to disclose, rather than secure, your personal data. It's worth taking the time to investigate those settings and make sure you're comfortable with them, just like it's worth checking whether you latched the windows before you left the house. When an application asks for "permission to send you push notifications and use your location data", take a moment to reflect on whether that's really what you want. Your camera and smartphone usually record the time and location in each photo you take, and when you share those photos, you may be publishing that data unless you specifically block it.

4. Find and use specific privacy-enhancing tools

There are many privacy-enhancing tools out there, especially for browsers. You can use them not just to protect specific areas of your digital footprint, but also to maintain your awareness and understanding of what service providers are looking at.[20]

With that layered approach in mind, here are some tips you can put into practice right away.

Manage cookies: Check what settings your browser(s) have for cookies; find your browser's "cookie store" and spend some time looking through it. Reflect on how many of the cookies in there have been set by sites you weren't even aware of visiting... and then see whether your browser allows you to block third-party cookies. While some browser settings help with this, many users have installed additional plug-ins to help them control tracking cookies.

Check your privacy settings: Erasing cookies only goes so far. Internet users must also take control of the information that they choose to share on any public service, especially explicitly open services such as social networks, blogs, and photo sharing sites. As with all private information, preventing exposure is simpler than the nearly-impossible task of erasing once let loose. Check what permissions apply to photos you upload, and consider expressing your preferences through mechanisms like Creative Commons licensing.

Understand the realities of data sharing: Once you've shared anything, in almost any context, you lose the ability to "un-share" it. And once you've visited a web site or created an account, you may lose the ability to erase your footprints. Even when a service promises privacy, the potential for unintentional data breach is always present.

20 Abine (www.abine.com) offers tools to add privacy to web browsing, including DoNotTrackMe, DeleteMe, and MaskMe. TrackMeNot (http://cs.nyu.edu/trackmenot/) is a browser extension to help protect web servers from surveillance and data profiling by introducing noise and obfuscation. Ghostery (www.ghostery.com) informs web users about the tracking information in their browser and on web pages, and block some types of tracking. Collusion (https://www.mozilla.org/en-US/collusion/) is an experimental add-on to Firefox to allow web users to see the third parties that are tracking them across the Internet. This list is just a starting point for readers; there are many more tools available.

Internet users also need to understand that a desire for privacy creates a conflict with many service providers, such as social networks. A social network is only useful when its participants share information broadly with each other. Social networks make it easy to share, and hard to keep things private. The interlocking mesh of social networks, photo sharing sites, blogs and micro-blogs, URL shorteners, and republishing services creates a significant barrier, keeping you from controlling your own information.

There's no quick fix here... but thinking seriously about the realities is a good step towards adjusting the value you place on privacy.

Give yourself the tools and motivation to make better decisions: By being mindful of the context for different Internet activities, such as "work," "personal," "social," "family," and so on, Internet users can increase control by using different software tools (such as different browsers) and different real-world objects (such as different payment cards and different smartphones) to create boundaries and limit the information that can be linked. While these techniques can be effective, they are also difficult to stick to.

The bottom line is this: good privacy is like healthy eating or good posture. We are most successful when we are motivated to switch from no (or occasional) privacy-enhancing behaviour, to privacy as something we do naturally and habitually; that means placing a value on privacy and personal data, which will sometimes over-ride our desire for convenience, or for a tempting but privacy-invasive app... just as a preference for healthy eating sometimes has to over-ride our desire for deep-fried lard.

It is sobering to reflect on this quotation from a former deputy director of the Federal Bureau of Investigation:

"Good luck trying to communicate in this world without leaving a digital exhaust—that's not going to happen."

~ Philip Mudd, June, 2013

Note that the metaphor that sprang naturally to his lips was that of exhaust... a toxic by-product of our preference for the convenience of the internal combustion engine.

Internet Society
Galerie Jean-Malbuisson, 15
CH-1204 Geneva
Switzerland
Tel: +41 22 807 1444
Fax: +41 22 807 1445
www.internetsociety.org

1775 Wiehle Ave.
Suite 201
Reston, VA 20190
USA
Tel: +1 703 439 2120
Fax: +1 703 326 9881
Email: info@isoc.org

bp-digital footprints-0114-en