PETER GASSNER
PROF. DR. GERHARD M. BUURMAN

ACTIVE IDENTITY

TRUST AND REPUTATION ON THE INTERNET

ZÜRCHER HOCHSCHULE DER KÜNSTE

ABSTRACT

I have researched ways for creating trustworthy interactions on the Internet. New online services like person-to-person lending platforms, that allow people to directly invest their money in strangers without an intermediary, require new forms of security and identification.

Based on work by PrimeLife and others, I envision a multilayered digital ecosystem that allows government grade security to coexist and be integrated with the global and open web. I'm taking the standpoint that a purely technical solution to security cannot exist and must always be complemented by social interactions. Therefore I'm presenting a tool that allows Internet users to actively control their online identities by integrating their activities and reputation from web platforms they participate in. This data is then visualized, acting both as a mirror of one's digital identity and a way to protect a user's privacy, minimizing unnecessary data exchange.

ABOUT THIS PUBLICATION

This is the Master's thesis of Peter Gassner, submitted to the Master of Arts program at the Zürcher Hochschule der Künste in the field of excellence "Interaction".

ACKNOWLEDGEMENTS

The ideas in this paper have been developed together with my mentor Prof. Dr. Gerhard M. Buurman. I received further valuable inputs from my lecturers Karmen Franinovic, Christian Weber, Raphael Perret, Max Rheiner and my classmate Judith Fehse. I would like to thank my partner Wilma Hunziker for her patient support.

Contents

1
Introduction

1.1 Context

As an interaction designer, I like to think about people and how they interact with the world. Through design, experiences can be created or improved to solve peoples' needs. In our department, the Interaction Design department at the Zürich University of the Arts, we feel that it is important to think about topics that go beyond our design knowledge, even if this means we have to be vague and maybe imprecise. Even though I am a layman in finance, game theory or cryptography, I think it is possible to create valuable input in these fields by means of bringing ideas together and envisioning a coherent story.

Many of the ideas in this thesis have been developed in close collaboration with my mentor, Prof. Dr. Gerhard M. Buurman. He established the Interaction Design course at our school in 2000 and co-founded the Swiss Design Institute for Finance and Banking[1] in 2007, an interuniversity competence center for design. It conducts theoretical and applied research to optimize the relationship between financial service providers and their customers.

[1] http://sdfb.ch

1.2 Research Interest

My thesis started out with the observation, that networked media allow us to rethink the way we handle money. I chose to look further into person-to-person lending because its adoption has been driven by innovative *online* platforms and made it possible for people to deal directly with each other without going through an intermediary like a bank. Person-to-person lending combines my main interest – networked media – with money, our society's central information medium; it brings people with vastly different demands together, creditors and debtors; and it relies on trust, which is a largely unresolved problem on the Internet.

It is especially this close interaction between debtor and lender that makes person-to-person lending interesting for my thesis. By making such a delicate topic as money the conversation starter between two strangers, their needs for security and trust are very different than if they just had to chat about recipes with each other.

2
Digital Ecosystems

I want to begin this thesis by describing the space where everything takes place: the Internet. Within comparatively few years, the Internet has gained an importance that has not been foreseen, and our reliance on it is growing everyday. While the medium was mostly used to consume information during the nineteen-nineties, this has changed towards a more bi-directional, conversational web in recent years; not so much due to technological change, but rather because people have found new ways of using the medium. I find the term "read/write web" especially useful to describe this change in thinking. Everyone is able to contribute to the many conversations, there is no need to know how to program. The web is no longer static, it's much more dynamic. Social applications like Facebook (created as a students-only network in 2004, but open to everyone since 2006) are a big driving force in bringing people to contribute content to the web. But it's also sensors and devices of all kinds that upload their data to the Internet ("of Things", as it is being called). With the Semantic Web on the horizon that tries to make all this content machine-readable, we will very likely see more creative uses of the medium, e. g. with even more sophisticated mashups (aggregations of data from various sources to create new meaning). A lot is happening.

Figure 2.1: Physical sensors of all kinds around the world can be accessed in realtime through Pachube, which in turn uses the Google Maps service as a display. `http://www.pachube.com` (14.5.2010)

I observe two trends. First, the web is no longer a place of insular platforms, instead, many different smaller services are connected and re-connected through application programming interfaces (APIs), drawing content from one platform, publishing it to another. These services often don't offer anything of their own, they just act as transformation functions on the content, redefining it, giving it new meaning. Their value arises from the network they're in. Second, the web is growing in new directions. It is no longer a mostly content-oriented medium, but is becoming an integral part of society, a place for social interaction, part of our culture.

The Internet is an open platform that allows for a lot of ideas to be tried out, with only the good ones staying around – often quickly replaced by better ones. It becomes obvious that this openness is the essential driver of digital evolution: one agent feeds on the other, which feeds on the other, which depends on still another one, and yet they cannot depend too much on each other because

the system is in constant flux. It is an ecosystem, ableit a digital one, and seeing it as such makes it obvious that we have to care for it as a society.

The term *digital business ecosystems* originated in 2002 as a way to think about how digital processes could benefit small and medium businesses[1]. In 2009 Briscoe and De Wilde[2] described *digital ecosystems* as the digital counterparts to biological ecosystems, highlighting their ability to self-organize, scale and automatically solve complex, dynamic problems.

[1] Francesco Nachira. Towards a Network of Digital Business Ecosystems Fostering the Local Development. May 2002

[2] Gerard Briscoe and Philippe De Wilde. Digital Ecosystems: Evolving Service-Oriented Architectures. Oct 2009

I'm interested in digital ecosystems not from a software engineering, but a social interaction perspective. It should be possible to identify oneself on the Internet in a trusted way, so that one can be held liable if something goes wrong. Also, it should be easy to build up reputation on the Internet and take it from one web platform to another. At last, the privacy of an individual should always be protected as good as possible. People should be in control of the information they share about themselves, and the kind of information they share should only include data that is really necessary. In my opinion, a digital ecosystem, where trusted information providers co-exist, but can be tightly integrated with ordinary web platforms, could solve these problems and create new possibilities for innovation.

Technically, the term *digital ecosystems* does not describe something new, but it provides a new framework for thinking about the various services on the Internet. I want to point out some core characteristics, that I find defining of digital ecosystems:

1. Digital ecosystems are *open* to everyone.
2. This allows for *evolution* to happen, because people can try new things and improve on existing ones.
3. Interfaces between service providers are as *simple* as possible and
4. they are *standardized* to foster exchange and optimize for re-use.
5. This allows systems to be *modular*, adding just what is needed, which in turn
6. makes them *scaleable* and *robust*.

2.1 A Solid Foundation

In the rest of this chapter, I will discuss the five layers shown in Figure 2.2. From top to bottom they are: *User, Interface, Platform, Control, Core Services*. In this section, I will talk about my reasoning behind the bottom layer, *Core Services*. In the following two sections, I'll go into detail about *Control* and *Platform*. The *Interface* layer describes the various devices (mobile or not) and interfaces a user has available to interact with a web platform. I will not discuss this layer further in this thesis.

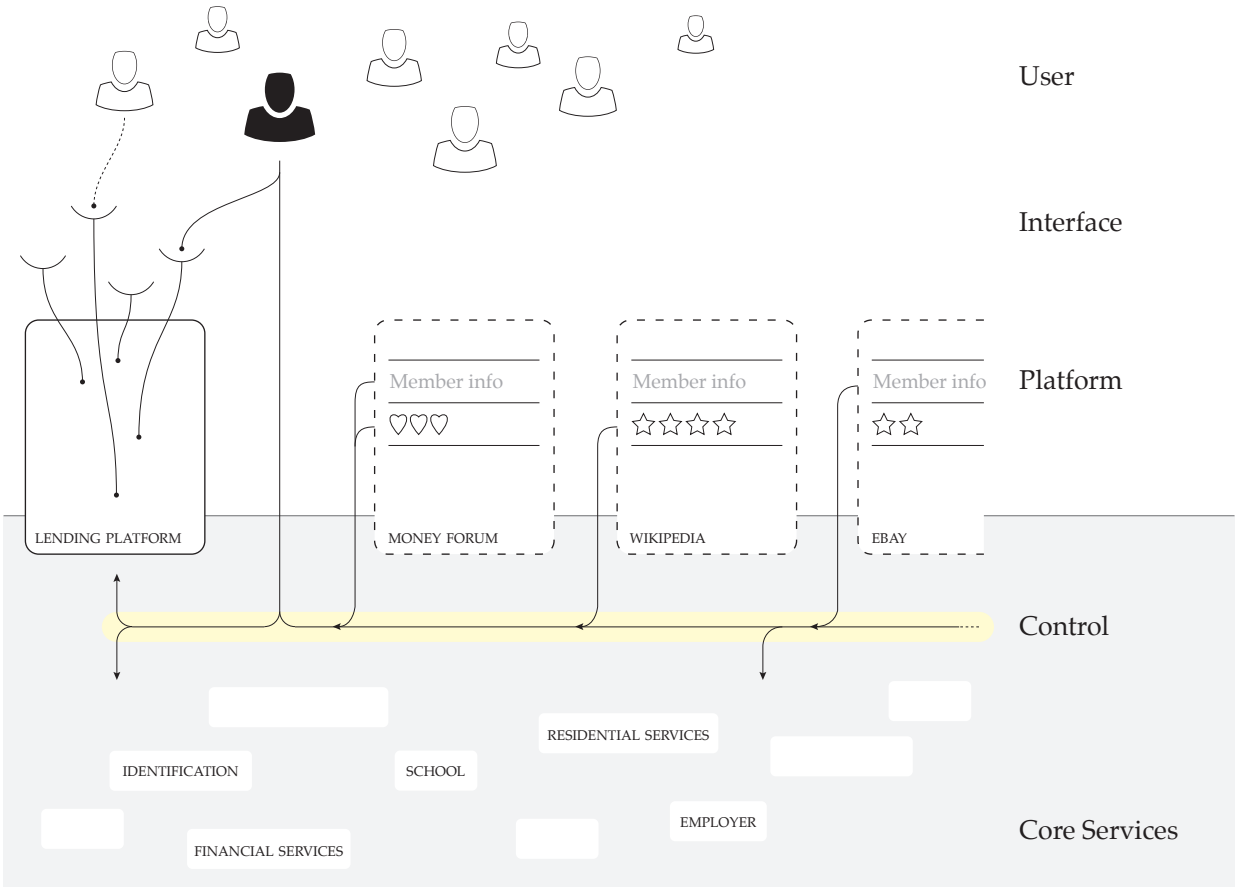

Figure 2.2: The five layers of a possible structure for digital ecosystems.

I don't know whether this system can work or if it is actually a good idea. However, it proved useful to think about the processes that take place and how concerns could be separated. And as I will try to show, the separation of concerns is essential to create an open *and* trusted system.

The foundation of the proposed system are *Core Services*. They are certified by global or national organizations and build a web of trust. An example of such a service is a national identity provider. Their job is to:

1. Provide trusted information (e. g. digital identification)
2. Transfer information securely between platforms (e. g. a tamper-proof channel)
3. Provide hooks to interact with untrusted platforms (e. g. person-to-person lending platforms)

Having these Core Services available, allows us to go metaphorically *deep* (see Figure 2.3). Instead of having to go to an office to fill out a form and present our passport, we could request this information from an identification service and hand it securely to the platform requesting this information. Or, if a person-to-person lending platform asks about our credit history, we could ask the Core Service of our bank to provide this information.

Consortiums like PRIME[3] and its follow up project PrimeLife have been formed to make such scenarios possible, although not yet on such a large scale as I imagine here.

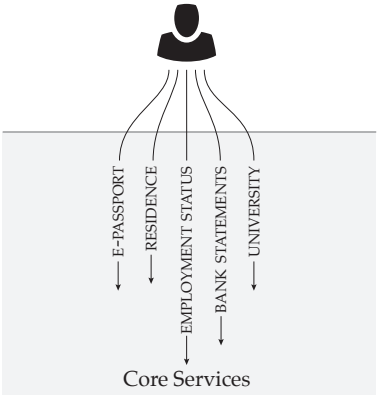

Figure 2.3: Core Services provide trusted information.

[3] R Leenes, J Schallaböck, and M Hansen. Prime White Paper V2. *PRIME Project*, 2007. URL `http://www.prime-project.eu/`

2.1.1 *Tokens Of Proof*

In his seminal book, Brands (2000) describes how cryptography can be used to create privacy, preserving digital certificates that can not be tampered with. He points out a danger of digital certificates:

> Unless drastic measures are taken, it will not take long before everyone is forced to communicate and transact in what will be the most pervasive electronic surveillance tool ever built. [...] The dossiers [contained within the digital certificates] can be compiled and linked without human intervention, can be dynamically updated in near real time, and will contain minute information about a person's financial situation, medical history and constitution, lifestyle, habits, preferences, movements, and so on.

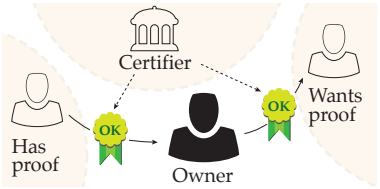

Figure 2.4: The information owner should be able to take a piece of information from one party to another without them knowing from each other. Through the use of certificates, the receiver can be sure that the information is valid.

Using such certificates, the control over their information is in the hands of the user. He can choose what aspects of the certificate to show to whom, while the reader can rest assured, that this information is indeed certified.

Figure 2.4 shows an application of this concept. The user wants to prove to a person-to-person lending platform that he is employed and earns over $2000 a month. He can then ask his employer for a certificate with a statement that he indeed earns more than $2000 (an exact amount is not necessary, more on this in Section 2.3.1). It should not include the name of himself, or any other way to draw conclusions on the owner of this certificate. Instead, the user can prove that this certificate is his own using his private key, that no one else has. Also, the certificate should not contain the name of the employer, as the user might not want to let the person-to-person

lending platform know *where* he works, just *that* he works. The reverse is true, also: the user doesn't want his employer to know that he is applying for a credit. To solve this dilemma, the employer can make use of a Core Service he is registered with (even though the employer himself isn't part of any Core Service) and have the information certified anonymously.

Even though it is a lot more complex and I have made assumptions that may be simplified, it becomes clear, that contexts can be kept apart, information can be transmitted anonymously, and only necessary information needs to be transmitted. If this certificate is stolen, it is basically useless to the thief.

The scenario I described made very specific use of Core Services in two ways: the certification process and the transport of these certificates. By having very few and well-defined touchpoints, it becomes easy for any web platform to make use of these services, which nurtures diversity.

2.2 *Nurturing Diversity*

As I have shown, openness is an essential driving force in digital ecosystem evolution. To keep the web open, there can't be any restrictions on what a web platform can or can't do. This means, that we can never trust any web platform, because they can do *anything*. This is why they have to be clearly seperated from the Core Services. This separation should be clearly visible to a user, e. g. by separating out Core Services into the browser or operating system like Mozilla proposed (Figure 2.5).

However, as I have shown in the previous section, having access to Core Services can lead to innovation and diversification. Instead of checking the backgrounds of people, a person-to-person lending platform could spend this money on innovative functionality. Also, the market would be more welcoming to newcomers who do not have a large customer base yet.

2.3 *Controlling Information Flow*

Because we cannot trust web platforms, they should not be allowed access to any sensitive information. Instead, the user should bring the information to them. This also means that web platforms cannot be allowed to communicate sensitive information with each other: the user is in control, and people, especially Generation Y-ers[4], like to be in control. But control also means work, and if something is too much work, people won't care anymore.

This problem will have to be solved through clever interfaces, where a choice will have to be made on where to store all these certificates. Will they be stored all in one place on the user's hard drive, similar to 1Password[5], or somewhere on the Internet? All in one place or distributed similar to what projects like OpenID and OAuth[6] do today?

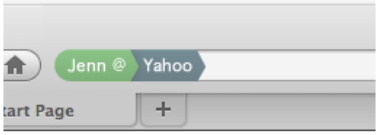

Figure 2.5: An example from Mozilla showing how identity functionality could be directly integrated into the browser (Raskin, 2009). `http://www.mozilla.com/en-US/firefox/accountmanager/`

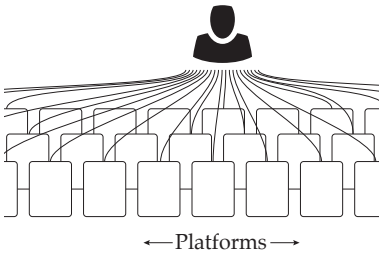

Figure 2.6: To foster innovation, the web has to be open to everyone. By providing Core Services, these platforms can provide certified information with little effort and cost.

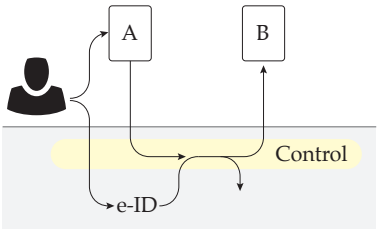

Figure 2.7: The user should be in control of his data at all times. This is the most essential layer.

[4] David Cox, Thomas L Kilgore, Tiffany Purdy, and Rekha Sampath. Catalysts for Change: The Implications of Gen-Y Consumers for Banks. *Deloitte*, Apr 2008

[5] `http://agilewebsolutions.com/products/1Password` (19.5.2010)

[6] OpenID is for authentication, OAuth for authorization. `http://openid.net`, `http://oauth.net` (14.5.2010)

2.3.1 *Data Minimization*

Control is not only about who will see the data, but also *what* they see. Today, this is often an all-or-nothing process: you either hand out all information about you, or none, in which case the person or platform who asked maybe reject to deal with you. Today, we violate our privacy in order to present an accurate, trustworthy picture of ourselves. We give out information for the whole Internet to see, often unknowingly that it will stay around. Some do this consciously, but many do so unconciously. Still others don't do much with the internet because of their fear of violating their privacy. What can be done?

Cameron et. al. introduced me to the concept of data minimization and derived claims in order to "protect privacy and avoid the unnecessary propagation of identity information".[7] I find this a very valid concept: a web platform does not need to know the exact date of our birth, it only needs to know whether we are old enough to use the offered services. So, a derived claim basically says "older than 18" instead of the original claim "born on day X in year Y".

[7] K Cameron, R Posch, and Kai Rannenberg. Proposal for a Common Identity Framework: A User-Centric Identity metasystem. *The Future of Identity in the Information Society*, May 2008

This is sufficient, because the platform asking this question only needs to make sure you are old enough. It becomes obvious, that a system based on derived claims will have to have some form of standardization in order to exchange claim information. But this won't be a problem, because, as I outlined at the start of this chapter, *standard* and *simple* protocols are at the heart of digital ecosystems.

But it's not only about security and identity management tools, but also about policy. If a website wants to confirm that a user is older than 18 years, they only need confirmation of this fact and not the birthdate of the user. Or if an employee wants to prove he earns more than $3000 a month, the party he is revealing this fact to doesn't need to know where he works and his employer doesn't need to know who he wants to tell this to. Without policies requiring care with data and encouraging privacy, everyone will try to gather as much data from a user as possible, creating loopholes for identity theft or similarly abusive misuses of identity information. This is definitely something we should strive to prevent, as I will show in Section 4.2.

3
Person-to-Person Lending

Person-to-person lending is a form of lending that has been rapidly gaining popularity since it was first introduced in 2005 by Zopa[1]. Lending money directly from one person to another is nothing new, in fact, this is how it all started hundreds of years ago. But because the creditor (the person giving the credit) runs the risk of losing his money if the debtor (the person receiving the credit) cannot pay back the money – be it through bad planning, unskillful use, or just bad luck –, the handing out of credits has been mostly in the hands of dedicated credit institutions and banks. They have, due to the many people saving their money in bank accounts, access to large reserves of money, they perform deep checks on debtors to calculate the risk of default, and they also have a lot of debtors who each pay interest that makes up for the loss of one or another debtor's money. It's an economy of scale.

[1] `http://zopa.co.uk` (19.5.2010)

The Internet has made two things possible that have previously prevented person-to-person lending from becoming popular. First, it scales: because creating specialized marketplaces has become so easy, people with similar interests find each other. This makes it possible to suddenly bring tens of thousands of people together instead of the dozen that were possible before due to geographic boundaries and the lack of communication possibilities. Second, the Internet empowers people to choose what interests them from a large set of options, no longer do they have to choose what is offered to them. In the case of lending, creditors are empowered to invest in debtors having goals they can relate with (although some do it for the profit, and others just want to leave their money in the bank).

Person-to-person lending platforms could be seen as just that: a couple of new websites. However, I want to argue that they are much more: instead of looking at these platforms individually, we should see them as part of a *digital lending ecosystem* (which is part of a digital money ecosystem, and so on). By taking this viewpoint, it becomes obvious, that person-to-person lending platforms are not the last word spoken on the topic, but rather the beginning of an evolutionary process that will lead to new ways of dealing with money. Unlike the read/write web, which I have introduced as being a more dynamic approach to content generation on the

Internet, person-to-person lending is about *social interaction*, not information. Although there are similarities to Facebook, person-to-person lending is a lot *riskier*, because people have the ability to exchange money with each other. Technical security alone does not necessarily help here, because people are in direct contact with each other. Instead, ways have to be found to create *trust* between people, something that has proven difficult on the Internet.

3.1 *Revisiting Money In A Networked World*

Figure 3.1: An excerpt of a poster I created to describe the research landscape. From left to right, topics merge and split to create new ones.

3.1.1 *The Internet Creates Global Marketplaces*

Since hundreds of years, marketplaces have allowed suppliers to find consumers and vice versa. Even though suppliers have to wander from town to town, people know where to find them: on the marketplace. While traditional marketplaces are regionally restricted, the Internet allows for global marektplaces to be created. Even though they are restricted by language barriers today, with the advancement of translation technology even this restricion might fall in the future.

Even more than traditional marketplaces, Internet marketplaces are driven by interests; instead of having a produce or a fashion market, cherry or shoe themed marketplaces are now possible. The creation of such specialized marketplaces is cheap and will more likely attract enough people because of the improved findability and the global customer base.

Figure 3.2: Like the fountain in a town's heart, digital marketplaces let people with similar interests meet each other.

Marketplaces are not only about exchanging goods but also about communities, people with similar interests. A community needs something to talk about, a "Social Object".[2] This is certainly not in short supply on a person-to-person lending platform. First of all, there is the credit that brings everyone together, but there are other interests: house owners, students, people with credit card debt, and so on. A community can bring these people together. Prosper has shown that there is an interest to form groups and that debtors belonging to a group are more likely to pay back on time. Wright, in a survey about Zopa, Prosper's competition, has shown that some people would welcome more community, while others would rather not want it.[3] Blaesi emphasizes, that community features are hardly used by current person-to-person lending platforms, even though this would be well supported by the medium.[4]

What I take away from this, is that digital marketplaces should revolve around a specific interest and yet be open to more than one community, allowing niches.

[2] Jyri Engeström. Why Some Social Network Services Work and Others Don't. Apr 2005. URL `http://www.zengestrom.com/blog/2005/04/why-some-social.html`

[3] Collette Wright and Michael K Hulme. Internet Based Social Lending: Past, Present and Future. Nov 2006

[4] Fabian Blaesi. Banking 2.0: Strukturelle Eigenschaften und Vertrauensbildung beim Social Lending. Dec 2008

3.1.2 Value Is Defined By People

I was taken by surprise when I researched money. I never really thought about it, what it is, where it comes from; I've just taken it for granted. I don't want to go too deep into the interesting story of money, but a short overview is necessary in order to define what is at the heart of person-to-person lending: money. Money has three core functions:

1. Money is a unit of account
2. a medium of exchange
3. a store of value

Money is a communication medium, one of the oldest and most important information systems known to us, writing has supposedly been invented for bookkeeping.[5] With money we can measure the value of things, an extremely powerful idea: instead of exchanging things directly – say, three liters of milk for six eggs –, the value of milk and eggs can be expressed with money. Because of this translation, it is no longer necessary to exchange goods with other goods, we can directly pay with money, even though money itself has no value at all, it is completely abstract. Having this abstract unit of account lets us store it. Konrad Paul Liessman has this to say about this:

[5] Bernard A Lietaer. Das Geld der Zukunft. *Riemann Verlag, München*, 2002

> "Geld ist [...] nicht nur erstarrte Substanz, sondern auch erstarrte Zeit. Jeder Geldschein, der nicht in diesem Moment ausgegeben wird, stellt einen Wechsel auf die Zukunft dar." (Liessmann, 2008)

Money is not only congealed substance, but also congealed time.

But who defines how much money something is worth? Who defines the value of things? The answer is: it depends. A physicist would set the value of a paper on atmosphere dynamics a lot

higher than a mother who has to buy dinner for her kids. He probably wouldn't know why the mother assigns such a high value to a baby-comforter. So, it depends on the context and the community, what something is worth.

Value is constantly defined and re-defined. This becomes apparent with the availability of alternative currencies. Alternative currencies can exist alongside national or multinational currency systems. They work in a similar way to money in that they let people exchange values freely. But they are most of the time bound locally.[6] An example of an alternative currency is the Bartercard[7]. It allows small companies to get a credit that they can pay back by offering their services to Bartercard customers in exchange for the same amount of Trade Dollars, which are worth the same as normal dollars, but can only be used withing the Bartercard system. In order for Bartercard to work, there have to be enough offers for customers to use their Trade Dollars on, a problem they solved smartly by giving credits to businesses like coffee shops that can actually offer a service to the community.

Figure 3.3: Frederic Vester boldly expressed the monetary value of a Bluethroat: two cents for the material, hundred and fifty euros for the work it does (eating bugs, soothing people, etc). (Vester, 1987)

[6] Bernard A Lietaer. Das Geld der Zukunft. *Riemann Verlag, München*, 2002

[7] Barbara Barkhausen. Tauschgeschäfte. *brand eins 03/07*, Feb 2007

Today, the Internet is a "money free" zone, mostly because it is really difficult to transfer money quickly and effortlessly, especially small amounts. But is it also *value free*? Certainly not. Even though there currently is no way to pay an individual contributor to Wikipedia, her work is of great value. It will certainly become easier over time to pay for work and information online, be it through online alternative currencies or simple payment systems that become available through other platforms in the digital ecosystem.

3.1.3 Knowledge Is Ubiquitous

There is another reason why handling of money on the Internet is different: knowledge. Like no other medium, the Internet makes knowledge available to everyone, anytime. Platforms like Wesabe or Mint[8] make good use of of this and create self-help communities around money. Let me give you an example.

[8] `http://wesabe.com`, `http://mint.com` (19.5.2010)

On Wesabe, a member asked how to get rid of his $4000 debt, of which he "can't tell mom". One community member shared this hint about changing one's behavior:

> Just a few months ago, I thought I knew exactly where my money was going, but learned [...] that I had no clue. Since then I've cut down on, for example, trips to the ATM. [...] Do you feel you know exactly where your money is going?

Another one suggested a book worth reading regarding the problems with "mom":

> Also....if you really think your mom has control issues, take a look at the book "Boundaries" by Henry Cloud. [...]

What I find interesting about these two answers, is that they are very personal and that they provide contextual information that does not necessarily have to do with money.

The reason I bring this up, is that I believe knowledge is safety. If someone can make an *informed decision* about a credit he's getting, he will run less risk of default. This also benefits the creditor, as his risk of losing his money sinks. Credit is powerful and easy to get at. This leads many debtors to underestimate its dangers.

To capture this line of thought, I created Figure 3.4 to show that being informed must also be actively communicated in order to gain the trust of a creditor. But few people – either because they don't know about it or because they're lazy – will seek information if they don't have to. This led me to introduce the concept of the *driving force*: lots of people will do anything for even the smallest rewards. This can be taken advantage of to make them do things that help them and that they may not do otherwise.

As I have shown in Section 2.3, the transport layer of digital ecosystems will allow these people to transfer the *reputation* they gain by informing themselves to other platforms, making the effort worthwile.

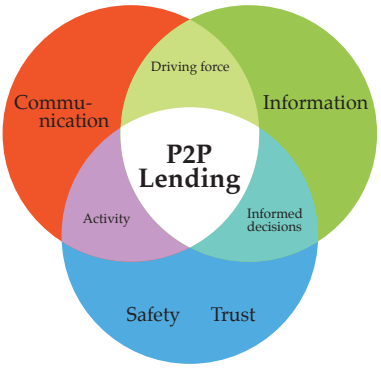

Figure 3.4: Person-to-person lending becomes safer if debtors are able to make informed decisions. If they can actively communicate this, they will be trustworthier.

3.2 *Everyone Can Be A Bank*

What person-to-person lending shows us, is that everyone can be a bank. We can, through the Internet, become creditors and debtors, maybe even at the same time. Because person-to-person lending works without a middleman, both creditor and debtor get better rates. But it also requires new mechanisms to distribute the risk of a debtor's default. The first person-to-person lending website, Zopa, appeared in 2005 and has mediated £25 million until 2008.[9] It was followed by several others, e. g. Prosper in the USA, Smava in Germany and Cashare in Switzerland.

When asked for the reasons why they participate in person-to-person lending, people cited the following reasons.[10]

50 % Profit

41 % Innovation

31 % Ethics

31 % Individuality

28 % Community

That people are interested in ethics, individuality and community makes it clear to me, that person-to-person lending is about people. Instead of letting a bank manage their money, interested lenders actively chose debtors who interest them. They want to know what happens with their money. If some people would rather leave their money in the bank, this is perfectly fine, the Internet is about *choice*.

There are two main ways these platforms mitigate risk: first, each applicant has to provide a credit history and will also be checked

[9] Fabian Blaesi. Banking 2.0: Strukturelle Eigenschaften und Vertrauensbildung beim Social Lending. Dec 2008

[10] Collette Wright and Michael K Hulme. Internet Based Social Lending: Past, Present and Future. Nov 2006

by the platform itself. Second, a creditor can never finance the whole credit of a debtor, his investment will be distributed among many debtors, resulting in shares of twenty to hundred dollars each.

The former is certainly a valid thing to do, not everyone should be allowed to receive a credit. But it does not conform to the "Internet way" at all, as it requires us to go to an office and show our papers, maybe even fax a form somewhere. This bureaucratic cruft is necessary because there is no better way at the moment, but the mechanisms I described in Chapter 2 could help to remove this necessity and thus open the system up for more innovation.

To conclude this section, I want to quickly introduce four kinds of platforms that show how the Internet can be used to handle money in an innovative way. The number and diversity of these platforms is rising rapidly.[11]

[11] weBank. Peer-to-Peer Finance Report. Mar 2009. URL `http://webank.org.uk/?page_id=73`

Person-to-Person lending The platform I have introduced in this chapter. A creditor can invest his money in debtors of his choice. The profit for a creditor is higher than usual, while the interest rates for debtors are lower since there is no middle-man. Popular platforms include Prosper.com and Zopa.co.uk.

Microcredit Entrepreneurs in developing countries often don't need large sums of money, to them fivehundred dollars is enough to e. g. buy seeds to start a farming business. This is about investing in interesting projects, there is no profit for the creditor, he will receive the same sum of money back. A popular example is Kiva.org.

Bartering Platforms like MachDuDas.de allow small projects like mowing the lawn or helping translate a text to be put online. There is no fixed price, applicants bid on the project, often including something special, e. g. "this is my favorite hobby, I do this all the time". So it's not only about money, but also about talent and hobbies.

Funding Kickstarter.com is a perfect example of a modern funding platform where people can support projects that interest them (e. g. a movie about an instrument). In return, they get small presents or parts of the project, depending on the amount they have donated.

3.3 Problems And Risks

I distinguish between security and safety. I define security to be measurable; the hardness of a lock can be measured and it is irresponsible to not use the hardest lock available. Safety, on the other hand, asks a more holistic question: is a person safe *in the end*. To answer this question, human – not directly measurable – factors have to be taken into account.

That said, I see three kinds of risks and problems that a person-to-person lending platform has to solve in order to gain the trust of its customers.

Technological risks No system is bullet proof and can be broken into. Still, it would be unwise to not employ the best *security* mechanisms available. Technological problems can be fixed, they are a known risk; in the perception of the public, however, they are rated graver than they really are, because the technology is new and not understandable – people have not the same grade of experience with technology as they have with other people. If something goes wrong on the Internet, this has a strong impact.

Credit risks Can the debtor pay back or will he default? This is a known risk and can be mitigated by statistics and risk estimation. As I have shown in Section 3.1.3, the Internet can be used to better educate debtors and creditors and thus reduce the risk of default.

Social risks People lie. A secure lock does not automatically lead to a *safe* environment. How can a creditor, who wants to invest in people asking for money to further their education, be sure they will really use it for this purpose? To them, lying would be easier than to justify their new car purchase.

This is very much a social problem, that humans have solved in two ways over time: reputation and trust. But these values are very difficult to communicate on the Internet. Michael Borter, CEO of the person-to-person lending platform Cashare[12] stated that "we know each customer personally, [...] we don't want people to cheat." This does not scale and it leaves the interpretation up to the lending platform. Using Core Services in a digital ecosystem could make use identity information available, on which decisions about the reputation and trustworthyness of a person can be made. I will discuss this in the next chapter, Chapter 4.

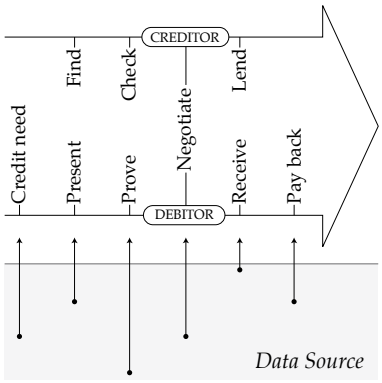

Figure 3.5: Over the course of a credit, there are many touchpoints between creditor and debtor. The more – and the more fact based – information is available, the easier it is to make decisions on both sides.

[12] Borter made this statement at Tweakfest 2010 in Zürich on 24.4.2010 `http://www.tweakfest.ch/de/?p=8541`

4
Identity

Figure 4.1: A passport is used to prove one's identity to others.

Identity is often used as an umbrella term and thus has many meanings depending on the context. I want to keep it simple: The term comes from the Latin word for "same" and means just that, two identical things are the same thing. A further distinction can be drawn between *qualitative* and *numerical* identity, stating that "Poodles and a Great Danes are qualitatively identical because they share the property of being a dog, [...] but two poodles [...] have greater qualitative identity", whereas "numerical identity requires absolute, or total, qualitative identity, and can only hold between a thing and itself".[1] In order to reduce confusion about identity, which is a central part of this thesis, I have phrased the following definition:

> The term "identity" refers to the numerical identity of a natural person. Thus, any single person has only one identity.

[1] Harold Noonan. Identity. *Stanford Encyclopedia of Philosophy*, Nov 2009. URL `http://plato.stanford.edu/entries/identity/`

Why this definition? Because the term is often used confusingly to describe the *identities* of a person. There is no such thing. The reason for this – in my eyes imprecise – usage stems from the observation that people do not *seem to be* the same depending on where we meet them. This is an important and correct realization, which unfortunately is hardly recognized by online environments: people have the need to strictly separate *contexts*, e. g. job and family, and need to be able to present different aspects of themselves, so called *personas*. However different these *manifestations of identity* may be, they ultimately point to a single person with a single identity.

4.1 One Identity, Many Roles

4.1.1 Personas

The word "person" is derived from the Latin "persona" (from PERSONÀR, to sound through) and was used to describe the masks that actors wore in ancient greek theater.[2]

[2] `http://www.etimo.it/?term=persona`

> It is probably no mere historical accident that the word person, in its first meaning, is a mask. It is rather a recognition of the fact that everyone is always and everywhere, more or less consciously, playing a role [...] It is in these roles that we know each other; it is in these roles that we know ourselves.[3]

[3] Robert Ezra Park. Race and Culture. *Glencoe, Ill.: The Free Press*, 1950

Persona is used to describe the part of a person's personality that is visible to others. This has been contrasted by Carl Gustav Jung with the use of the term "anima", the part of the human psyche that is directed inwards. This distinction between an internal and an external self has been "constructed in various ways", as Boyd points out.[4] She used the terms *internal* and *social identity*.

The recognition that people play many different roles everyday has also been the study of the sociologist Erving Goffman in his aptly named book "The Presentation Of Self In Everyday Life".[5] He talks of the many *theatrical performances* everyone of us is giving each day. As an example of this, he describes how a girl in a girl's dormitory tries to impress her roommates by arranging phone calls in such a way, that everyone can hear that she is being paged.

[4] danah boyd. Faceted Id/entity. Aug 2001

[5] Erving Goffman. *The Presentation of Self in Everyday Life*. 1959

4.1.2 Contexts

Goffman points out that "audience segregation" occurs: individuals ensure that those before whom they play one role will not be the same individuals before whom they play a different role in another setting. They separate *contexts*.

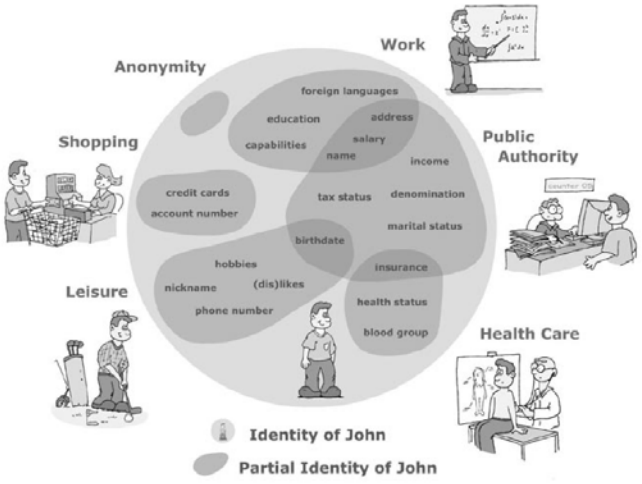

Figure 4.2: People present different facets of their identity depending on context. (Hansen, 2007)

Playing roles comes very naturally to us and we are aware of this in others, too. When we meet someone face-to-face, we can use our instincts to "see through" someone, to detect the slightest inconsistencies in their acting, as Goffman puts it. On the Internet, our possibilities for expression, control and observation are markedly different. Because a lot of what we do online is being stored somewhere and can be accessed through search engines, we have a lot less control over our audience. Boyd gives this example:

> When one presents oneself at a pub, most likely they do not expect that their presentation will reappear at work to be considered out-of-context. (danah boyd, 2001)

By aggregating online data about someone, we get to see a lot of information that is out-of-context. The conclusions that can be

drawn in such a case are of little worth, as the reason (context) for a certain behavior is hidden from us and was probably not meant for us to see.

4.1.3 *Identity Management: We've Been Doing It All Along*

Identity management sounds complicated, and unfortunately, it is. While it has been shown that people are very adept in projecting a specific persona to others, this is something that is incredibly difficult to achieve online. Identity management is an unnatural solution for an unnatural environment.

> In daily interactions, people are aware of their presentation: they know what they are wearing, they have a sense of their facial expressions, and they can easily comprehend the reactions presented by others. (danah boyd, 2001)

Online, however, "the lack of embodiment makes it difficult to present oneself and to perceive the presentation of others", as Boyd points out.

Identity management tools like the Higgins I-Card try to take contexts into account and provide users with identity cards, whose contents the users can define themselves. Depending on context, the user can then use one or another card to present himself.

Figure 4.3: The Higgins Project's I-Card. (The Eclipse Foundation, 2009)

Despite such tools, identity management is still in its infancy, mostly because users are not even aware of the information they spread online. This is part of the reason for creating the identity visualization I will show in Chapter 5, which should make users aware about what other people see of their identity online.

4.2 *The Importance Of Privacy*

Prof. Geoffrey P. Stone, in an interview on the topic of privacy, wants us to think about the following statement: "Imagine a world where 24/7 everything you do is recorded, available to be looked up by anyone. That would undoubtedly change one's behavior in all sorts of ways. [...] You can control people if you know something they don't want anyone to know."[6]

As I have shown in the previous sections, we lose control over context when we give up our privacy. Unlike identity management, privacy management is not about communicating who we are, but about what information others know about us, without us being aware of it. It is just as difficult to do:

> We now understand that privacy management cannot be addressed solely or even largely by a static set of preferences that determine how a user's information can and cannot be shared. Rather, privacy management is a fluid, organic process in which users are constantly refining their choices based on any number of contextual facets.[7]

A very important reason for privacy is that, on the Internet, everything is public, and we often unknowingly share information

[6] Prof. Geoffrey P. Stone, University of Chicago, `http://privacyrevolution.org` (2.5.2010)

[7] J Goecks and Elizabeth Mynatt. Using Social Methods to Support Privacy Management. *Security & Usability: Designing Secure Systems That People Can Use*, 2005

we would rather not share. When we sit in a restaurant and talk to our vis-à-vis, our neighbors can listen to whatever we're saying. This is not a problem for us, because they don't know us, and thus – more importantly – can't *leak information* to parts of our social circle that should not know about the contents of our little chat. This is an entirely different situation on the Internet.

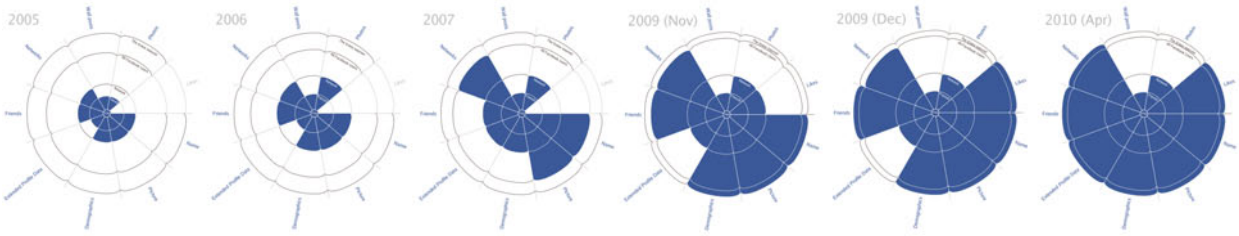

Figure 4.4: The Evolution of privacy default settings on Facebook. The center is private, the outmost ring completely public. The data was derived from the Facebook Terms of Service over the years. Matt McKeon, `http://mattmckeon.com/facebook-privacy` (14.5.2010)

Unfortunately, the most successful social network, Facebook, encourages its members to disclose all information they put online (see Figure 4.4). Websites like YourOpenbook.org show how absurd this information looks when taken out of context. Initiatives like these try to raise awareness, but danah boyd reminds us, that

> Making something that is public more public is a violation of privacy.[8]

[8] danah boyd. Making Sense of Privacy and Publicity. *SXSW, Austin, Texas*, Mar 2010. URL `http://www.danah.org/papers/talks/2010/SXSW2010.html`

By the time a piece of information is on the Internet, it will probably stay there forever – the Internet does not forget. Just think about your email history that Google has access to, the Internet Wayback machine[9] or these stupid comments you left ten years ago in a forum and that still can be connected to you. And these often are connections you would have never thought to make! Professional identity resolution systems will find this information and maybe prevent you from getting a job.

[9] `archive.org` (19.5.2010)

Having all these GPS-enabled mobile phones available makes us even leave traces of where we are and where we have been – for the whole world to read. An interesting anecdote from a Wired journalist tells us, how he saw a woman take a photo in a park using her iPhone. In the evening, he went home to look on Flickr.com, what pictures have been taken at this location today. Of course he found the woman including photos of her apartment, her address, and so on.[10] This may be a harmless example, but websites like PleaseRobMe.com try to make us aware, that when we share information about our whereabouts, that this could very well be exploited, e. g. by robbing our – obviously – empty house.

[10] `http://www.wired.com/print/gadgets/wireless/magazine/17-02/lp_guineapig` (16.5.2010)

At a conference, a person told a story about how he was shown data that a big supermarket chain was able to gather from its customers. Apparently, it was very detailed, and when he asked how they got this data, the answer was basically "we offered them a coupon for a doughnut and a cup of coffee."[11] People are not aware how valuable their data is to others.

[11] J D Lasica. Identity in the Age of Cloud Computing. *The Aspen Institute, Washington, DC*, Apr 2009

Private data can also be stolen, which happens all too often, and reveals credit card numbers or login credentials to the thieves. Sometimes, security leaks also have more direct social implications, as in a security leak on Facebook that allowed you to see the private conversations of your friends.[12]

The examples are many, but still, we often neglect privacy. Sometimes we might even be led to believe that surveillance can be a good thing, as in the case of a murder conviction, where Google search results "proved" the person was guilty.[13] Whether this is indeed a good thing or not is not the topic of this thesis. However, we should be aware of privacy risks when designing applications for the Internet.

By consequently applying data minimization techniques as described in Section 2.3.1, many of these privacy risks could be minimized or completely prevented.

[12] urlhttp://eu.techcrunch.com/2010/05/05/video-major-facebook-security-hole-lets-you-view-your-friends-live-chats (10.5.2010)

[13] `http://news.cnet.com/8301-13578_3-10150669-38.html` (10.5.2010)

4.3 *Reputation Is The Currency Of The Internet*

In order to create a successful digital ecosystem, people need to be able to trust each other, "we have to overcome the barriers around trust, reliability control and security".[14] But what is trust?

According to a definition by Luhmann,[15] "trust is a mechanism to reduce social complexity". Normally, when we meet a stranger, we have to be distrustful and constantly evaluate, whether the actions of this person reflect our expectations. If, over time, our expectations are fulfilled constantly, we can start to trust this person, meaning we don't have to constantly evaluate her actions anymore. Thus, the relationship has become less complex.

[14] J D Lasica. Identity in the Age of Cloud Computing. *The Aspen Institute, Washington, DC*, Apr 2009

[15] Niklas Luhmann. Vertrauen: ein Mechanismus der Reduktion Sozialer Komplexität. *Lucius und Lucius, Stuttgart*, (4), May 1968

Because trust takes so long to build up, we often rely on intuition, making a quick – and possibly risky – decision to trust a stranger to a certain degree. On the Internet, we rarely have any information available to make intiuitive decisions. Also, the information could be fake, luring us into trusting a fraudster. Reputation can help us out.

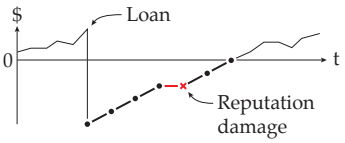

Figure 4.5: Reputation is hard to build up and easy to damage. In the USA many students get a loan to improve their credit score by proving they can pay back.

Figure 4.6: On Ebay vendors cannot afford to lose reputation by selling bad products or being unfriendly, reputation is their currency.

Reputation is a social metric for predicting a person's future behavior. If, over a large span of time, this person has always acted to the observer's prediction, he can put more (or less, if the reputation is bad) trust into this person. The person gets a good reputation. It takes a very long time to gain a good reputation and only very few incidents to lose it. In the context of lending, someone has a high reputation when they always pay back their credit on time.

There are many ways to gain some kind of reputation on every online platform: on Facebook you're good when your friends are good, on Ebay vendors get rated for the quality of the products they sell, forum users that help are more respected, bloggers that write well can influence others, programmers get a voice when they contribute good code to a community, etc.

However, these reputation systems are very insular, the reputation gained on one platform cannot be used on another. Through

Core Services of the digital ecosystem, we could connect these reputation sources between platforms in a tamperproof, anonymous and reliable way (Figure 4.7).

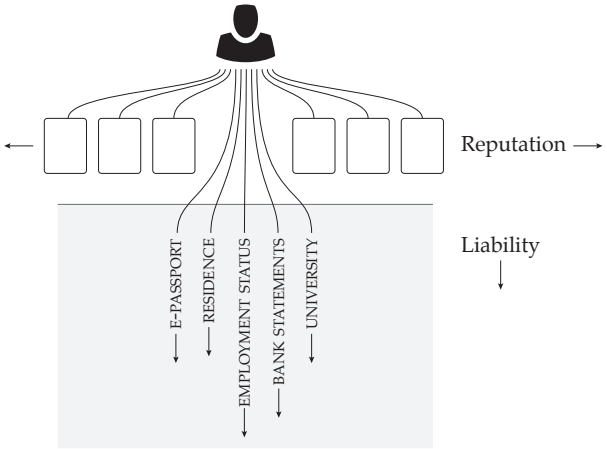

Figure 4.7: We can only trust someone if this person can be kept liable. If this person can present us proof of a good reputation, it is easier to trust.

On the Internet, reputation is a currency. If a person can present us proof of a good reputation, this basically saves both parties time. Trust does not have to be built up over time, but we can rely on the person's reputation instead (never completely, of course). We can also trust more, if the person can be kept liable for her actions, and if the person can provide us with certified relevant information, if needed.

Because reputation tells a lot about the behavior of a person, it is very important to handle this *delicate data* with care and keep it as anonymous as possible. As Prof. Dr. Urs Fischbacher[16] pointed out, it would be too easy to find and exploit good-natured people that are quick to believe anything.

Making reputation portable could have other interesting consequences. Today, we create a new user account on every website we visit. This is often abused by malicious users who can easily create throw away accounts and discard of them when they have been marked as fraudulent. But portability is not only useful to prevent fraud, it is also an estimation of the "work" a user does when he, for example, contributes to an article on Wikipedia. If reputation is a currency, people will care more about it and become aware of it on the Internet.

[16] Prof. Fischbacher is Chair of Applied Research in Economics at the University of Konstanz and expert in Game Theory. I invited him to discuss person-to-person lending on the Internet on 27.11.2009.

5
Active Identity: An Identity Visualization Prototype

To find out more about how the manifestations of a person's identity on the Internet could be used to create trustworthy relationships between strangers, I created an identity visualization. This is not an attempt to create *the* identity visualization, but rather a prototype to express and develop my thoughts, to find out whether this idea can work. This is a very delicate topic, as it is treading the line between showing too much, showing the wrong things, distorting reality, generalizing, lumping unrelated people together, and so on. Nonetheless we have to think about this, because the current practice of sharing *everything* or nothing about ourselves for lack of better options has to be changed.

Why a visualization? I had three goals: communicating the data about a person (mostly numbers) in a readable and comparable form; making this data available in an anonymous fashion to guard the privacy of the visualized person; and last but not least to motivate people to see what their – and the identity of others – looks like, making them aware of their online behavior. By showing "patterns of conventional use and the deviations from them", security becomes usable.[1]

I call it the *Active Identity* to emphasize my strongest point: it does not automatically aggregate data about people, people have to *actively* add it themselves. This means that everyone's visualization will be empty at first. This is the exact opposite of the excellent project "Personas" by Aaron Zinman[2] which wants to show you how the Internet sees you, with all the mistakes and guesses the computer makes. This critical work shows an unfortunate aspect of digital life: data gets mixed up and is attributed wrongly. It also finds older data or data that you don't want to show up and thus constructs an inaccurrate and inappropriate image of you. This is a reflective work, but similar software is being used daily to, for example, screen applicants to a job.

Instead of aggregating random data from unknown sources, I want to do the opposite: the user should be able to control the data sources that the visualization consumes information from. He should not, however, be able to manipulate the *data* that goes into the visualization, it should be based on facts.

[1] P DiGioia and Paul Dourish. Social Navigation as a Model for Usable Security. *Proceedings of the 2005 Symposium on Usable Privacy and Security*, 2005

[2] Personas – How Does The Internet See *You*? `http://personas.media.mit.edu` (13.5.2010)

5.1 *Facts, Not Fiction*

A fact expresses a truth about something, but as we've seen, people play roles, they act, so can *anything* we know about a person be a *fact*? Certainly, physical properties are indisputable (blue eyes), but they are not important online. It is the manifestations of a person's identity that we're interested in, e. g. Twitter status updates, Flickr photos, Facebook likes, etc. If such a manifestation can be originally attributed to a person, it is a fact: a truth about this person.

With this definition, I can say that my identity visualization is based on facts because it does not contain made up or automatically aggregated data. Instead, all the data that the visualization is based on is *actively* contributed and controlled by the user. How can this work without the user trying to improve his own visualization by providing forged data? The answer has two parts.

First, I'm relying on the solutions described in Chapter 2, that are partly hypothetical and partly available today in the form of OpenID and OAuth (see Section 2.3). The prototype in Section 5.5 uses OAuth in order to gain access to the user's data. The access to this data can only be granted by the user. Because my application consumes the data directly from the data provider, the user cannot tamper with the data. This solution works the way it should, but has the disadvantage, that the exchanging platforms know of each other, which should not be the case as I have described in Section 2.3. Also, if the user grants my application access, it will be able to see a lot of personal information. This is entirely unnecessary and could be avoided by using derived claims (see Section 2.3.1).

Second, the user is in control of what data enters the system, so if he has something to hide, he doesn't add this data. *This is perfectly fine, and it should be*. The privacy of the user is essential and has to be respected at all times (this is protected by law in many countries but is often disregarded online). This is the main reason why the user should be in control. But if the user doesn't add enough data to prove his reputation, this will of course be reflected in the visualization. Also, if the user suddenly removes a data source, this has consequences on the visualization: the data will be gone, but people will see that there is something missing and are able to decide what to make of this.

5.2 *Abstraction*

A successful abstraction tells everything we need to know, no more, no less. As Figure 5.1 shows perfectly, we don't need to see every metaphorical wrinkle of a person's identity, a good abstraction is able to communicate what we need to know, no more. It is also easier to create groups of identities, based on similarities in one or another parameter, making them comparable and understandable.

The visualization does not make use of portrait photographs,

Figure 5.1: Even an abstract face can still be recognized as a face. Depending on the grade of abstraction, a specific face or a more general group of faces can be depicted. (McCloud, 1993)

something we're used to from passports and other means of identification. The reason for this is, that we are easily deceived by good looking photograph. In fact, it has been shown, that beautiful politicians are more successful than their less good looking competitors.[3] This is often exploited in online forums, because it is easy to make a fake but beautiful photograph your avatar. The solutions discussed in Chapter 2 could of course be applied to ask for a certified photograph of the person. But this is a) complex, b) maybe not something a user wants to show to everyone, and c) photos are irrelevant on the Internet, as you never meet the user in person. Instead of beauty, a visual abstraction lets us introduce other, more relevant facts about a person's identity.

[3] N Berggren, H Jordahl, and P Poutvaara. The Looks of a Winner: Beauty and Electoral Success. *Journal of Public Economics*, Jan 2009

5.3 *Parametrization*

What does it mean when someone has many friends on Facebook? That he is especially friendly? That he is respected? Or that he is addicted to collecting friends online? It is obvious that raw numbers can't express who someone is, the data has to be interpreted with context, intent and experience. Although algorithms can make precise predictions when there is enough data available – something that is easier to find on the Internet than anywhere else –, the final judgement should be left to the person looking at the data.

I am not a statistician, nor do I know any of the data sources I used for my prototype (see Table 5.1) good enough to be able to create good mappings from the data of one provider to the slightly different data of another one (whether this is actually possible is questionable). However, for my purpose – to explore what a fact-based identity visualization could look like –, the goal is not to find a perfect algorithm, but rather to find out about possible parameters to base a visual form on. What matters most to me, is that the parameters are consistent, comparable and are based on real data.

CATEGORY	TWITTER	FACEBOOK	LINKEDIN
Avatar	full name, member since, personal website, description	full name, personal website, description	full name, personal website, short and long description
Status	total status updates, daily average[a], recent average[a], percentage of links[b]	average status updates[c], recent average[c], number of events participating	–
Relationship	followers, following, percentage of conversation[b]	friends, groups	friends, recommendations
Geography	location, time zone	time zone	location

Table 5.1: [a]Data from `http://followcost.com`, the recent average spans the last 100 updates. [b]Data from `http://mrtweet.com`. [c]Interpolated from profile feed.

Working with real data presented challenges in deciding on what can be expressed, how to get enough real data, and finding ways for handling the unique combinations of data providers that each user could compile individually. This markedly influenced the

resulting visualization and showed me how difficult it is to work with inhomogeneous data sets.

I came up with three parameters that I could get reasonable data for and that I assumed would be different between individuals: *Activity* to describe how active the person is on the Internet, and – more importantly –, to point out, whether there are substantial changes in activity over time; *Social Network* to describe how many contacts the person has, and whether he participates in any events or groups; and *Reputation* to describe the quality of the social network (web of trust) and the overall online appearance regarding consistency and reliability?

"Reputation" is an especially ventured parameter: it's a big word with deep meaning (see Section 4.3) that can hardly be put into a single number. However, I wanted to use this term exactly for its weight, in order to evaluate how it compares to the other two in peoples' opinion (see Section 5.5.2).

To use these parameters in my visualization, I needed to normalize them to values between zero and one. For this I first created accumulation functions for each parameter that took the values in Table 5.1 which I deemed applicable and then summed them up putting a weight on each value (Table 5.2).

Activity	=	status updates
Social Network	=	conversation; friends; followers; groups; following; events
Reputation	=	number of recommendations[a]; consistency of name, website and time zone; friends and follower/following ratio[b]

Table 5.2: First named values have more importance. [a] Available on LinkedIn. [b] The *quality* of the social network would be a better indicator, i. e. the Web of Trust.

To normalize these parameters to values between zero and one, I started out by using the average and maximum values. This turned out to be a problem, because people with values far above average made everyone look bad in comparison. So I started to use the *median* and the 10^{th} and 90^{th} *percentile*, because this puts less importance on outliers and is a better fit when working with unknown boundaries. Of course, this is something that would have to be constantly analyzed and re-evaluated to apply the correct statistical method.

All in all, I can say that preparing the data for the visualization was by far the most difficult part. I would like the data to be as transparent as possible, yet this is really difficult to do, especially if a lot of data points get condensed to a small number of parameters. My statistics probably lie. Still, the goal of getting consistent and comparable data points has been achieved and they can now be visualized.

5.4 *Visualization*

There is no perfect and unmistakeable visual form to express something as complex and diverse as identity. It starts with cultural differences in the interpretation of color and form (which probably could be solved if the visualization were localized, similar to language translation), but goes much deeper when we try to visualize the vast amount of data that makes up an identity and often isn't visual at all. Nonetheless, I do believe that it is possible to visualize *certain aspects* of an identity and that doing so can increase privacy, improve readability and reduce fraud through identity theft.

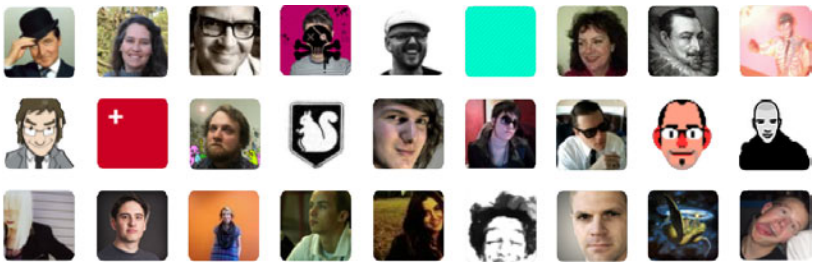

Figure 5.2: Globally Recognized Avatars are being used to create a consistent appearance across platforms like Twitter, Facebook, etc. `http://gravatar.com` (17.5.2010)

Figure 5.2 shows a collection of Gravatars, Globally Recognized Avatars. These are typically used on platforms like Twitter, Facebook, etc. to communicate a coherent and personal image of yourself. As we can see, portrait photographs are used a lot, some are illustrated, others painted on, and still others are obviously not the portrait of the user, but of a person this user likes (e. g. top left). These Gravatars are a great way of communicating personality and interests, but they do not work on their own. We have no way of identifying whether this Gravatar communicates something true or whether we've fallen into a sympathy trap. It is also trivial to take the Gravatar of, say, Cory Doctorow (third image) and add this to a comment on a random blog. Nobody could really tell, whether this is Cory speaking or whether it is an impostor. This is why I want to experiment with a fact based identity, as I have described in Section 5.1.

Figure 5.3: `http://www.digital-identities.com` (10.1.2010)

Another beautiful approach to identity visualization has been shown in the project "Identität – The 'Gestalt' of digital identity" where the authors used four parameters to represent the online identities of several persons: *interests* (represented by tags on delicious and twitter), *communication behavior* (ratio of dialogues and monologues on twitter), *activity* (in online communities), and the *age of the digital existence*. This project's focus lies on the individual form, whereas I want to create readable and comparable shapes that can be put side by side with hundreds of others, so I'm looking for something more ordered.

Ware[4] describes four features that we are trained to recognize: *form*, *color*, *motion* and *spatial position*. I wanted to take advantage of this and started with creating a distinctive shape, with the main

[4] Colin Ware. Information Visualization: Perception for Design. *Morgan Kaufmann*, 2000

goal being that it can be displayed very small and still communicate its most important properties (and defects). This way, many hundreds or thousands of identities can be displayed simultaneously and distinguished through the visual processing power of the human brain.

Beauty, even though I dismissed it above because it can lead to prejudice, can also make a visualization more readable: our eyes are trained to recognize symmetry and good proportions. Schmidhuber (1998) found that the shorter an algorithmic description of a face, the more beautiful it appears. So I wanted to look for a form that looks beautiful if it represents good values.

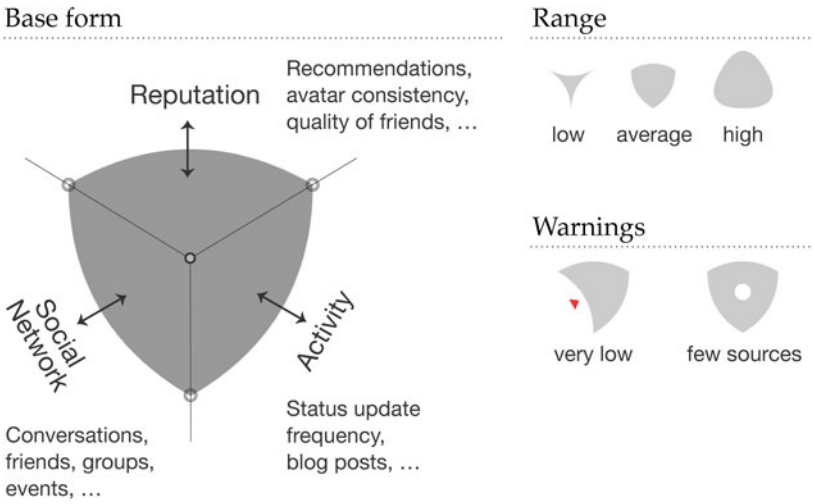

Figure 5.4: The basic form has three dimensions. Depending on whether the respective values are lower or higher than average, the form looks well-fed or meager.

I've described the three main parameters for my visualization in Section 5.3. They are *reputation*, *activity*, and *social network*. By assigning these parameters to the three sides of a triangle standing on its tip, I was able to create a simple shape that shows the quality of each side's data understandably by being well-fed or meager, respectively (see Figure 5.4). This metaphor does of course express a judgement, namely that if a parameter has a low value, the shape looks meager and thus is bad. Despite this, I want the shape to be as *neutral* as possible, so that observers can decide for themselves, whether they accept a low activity score as long as the person has a good social network, etc.

Taking this basic shape, we can now create a typological grouping as shown in Figure 5.5. Even though every shape is different, we can clearly differentiate between the different groups (rows). This means that we don't have to look at each shape individually to be able to read it, which makes it easier to understand. We could say that instead of looking at single letters, we can look at words. I want to point out that the size of the basic triangle is the same in each shape, and still the area and the sizes vary markedly. This is an additional aid in reading the shape, so this dimension should not be parametrized.

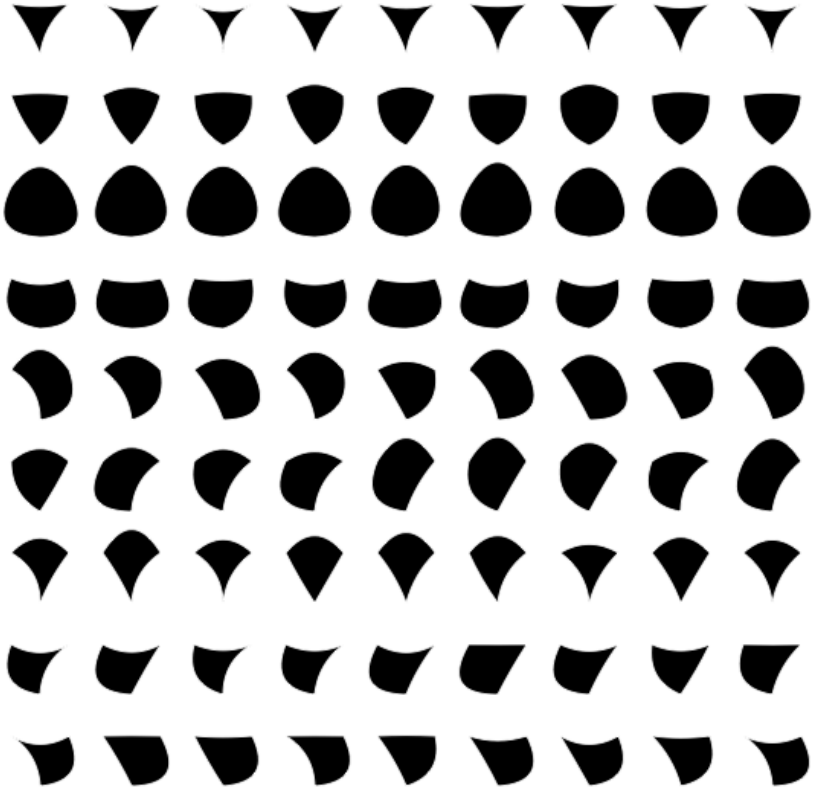

Figure 5.5: A typological grouping of possible identities. The second row shows identities with more or less average values in every dimension.

While people should be able to interpret the basic shape neutrally, I wanted to communicate anomalies and possible dangers very clearly. I identified the following three dangers and made them visually stand out:

1. Recent change (color) – In order to prevent abuse, e. g. by creating fake user data, I wanted to clearly communicate anomalies in user activity. If someone suddenly has a lot of data, there is a possibility of fraud.

2. Few sources (punch hole) – If someone has added very few sources, there is a danger that these might be fake.

3. Extremely low value (red triangle) – Values way below average are marked.

The color palette is chosen with brightness in mind: the most important color is the darkest and will thus stand out more. Also, I chose green as an indicator for "no problem" because this is widely understood in our culture. Red indicates a possible danger that careful attention should be paid to, because the user has become very active recently, the reason for which could be that someone wanted to create the impression of being active or that the account has only recently been created. A fall in activity is less suspicious but still something to pay attention to, so I colored it blue.

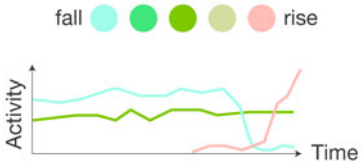

Figure 5.6: Time can reveal sudden changes in activity which may point out suspicious behavior.

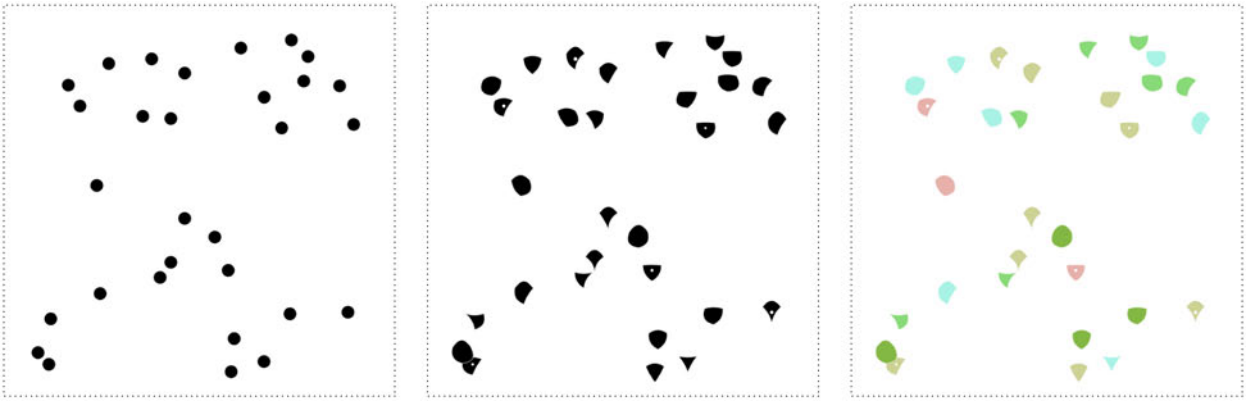

Figure 5.7: Details should be presented, as the human visual system can easily cope with them. As a next layer of detail, text should be introduced.

A stated goal was to make the visualization readable even if many identities are displayed simulatneously. I want to show how my solution works on the basis of Figure 5.7. Going from left to right, I introduced more and more detail. Even though it is commonly said that too much detail is confusing, I am always surprised by how good we are at reading complex maps. The eye should be given as much information as possible, as it will be able to sort out what is important and what is not. It can only do this if information is available. Colin Ware has this to say on the subject:

> Why should we be interested in visualization? Because the human visual system is a pattern seeker of enormous power and subtlety. The eye and the visual cortex of the brain form a massively parallel processor that provides the highest-bandwidth channel into human cognitive centers. At higher levels of processing, perception and cognition are closely interrelated, which is the reason why the words "understanding" and "seeing" are synonymous.[5]

[5] Colin Ware. Information Visualization: Perception for Design. *Morgan Kaufmann*, 2000

If we think about typology again, it becomes apparent that even though the figure in the middle introduces fourty shapes that are never the same, we can visually group them. We can ask questions like "where are the shapes with good reputation?" quite easily, something we're not at all able to do with the figure on the left. If we go one step further and add color, we can clearly see, that there are three identities that show suspicious activity patterns (red). In addition to this, the visualization of two of those is based on little data, so we should carefully look into them before we interact with them.

Something else I want to point out is the spatial grouping, which is a powerful means of showing relationships. Text should be introduced to label these groups, as it is the most expressive medium to convey meaning.

As you can see, I have not labelled anything in this figure. The reason for this is context: depending on the situation, the relationships between the individual identities can vary. Maybe it's not the relationships between individuals we want to show but their relationship to certain topics. Or something else completely. Spatial

position should be re-evaluated on a case-by-case basis to find out what should or can be communicated best with this parameter.

A decision has to be made on whether the basic dimensions of the identity shape should be changed on a case-by-case basis. There are good reasons for this, e. g. because a specific parameter of the shape doesn't apply to the context it is used in. On the other hand, the shape itself becomes less readable between different contexts, because the dimensions say different things. This is something that will have to be tried out.

What I tried to make clear in this section, is that an identity visualization can be made readable. Visual abstractions are no more than cute images if they cannot be read. There needs to be a common denominator, a basic shape, in order to communicate meaning through deviations thereof.

5.5 *Working Prototype*

I made a working prototype in order to try out, improve and communicate my ideas about identity visualization. I created this prototype using the Javascript and HTML 5 programming languages, because they can be executed natively in any modern web browser and are open technologies, supporting my argument for an open web.

Figure 5.8: Using OAuth, users can connect their identity information from various sources to my application.

To experiment with the ideas of connecting to Core Services of a digital ecosystem, I wanted the prototype to be based on real data. To get at this data, I had two options: I could either let a user enter their public profiles on Facebook, Twitter and LinkedIn, or connect to their profiles via the APIs that these providers offer. I chose the latter option because it resembles connecting to Core Services.

What I found, was that these APIs offer *way* too much access to a user's data. And, except for LinkedIn, this access is not restricted to a certain date. This was a very educational experience, raising my awareness for data minimization even more.

Besides using OAuth to authorize my application to connect to these platforms, I integrated OpenId. OpenId is not about granting access, but rather about confirming the existence of a user and the validity of his credentials (authentication). Where a user would traditionally have registered an account on my platform, OpenId makes it possible to have a central account, with which a user can login to any website that supports it. This helps a lot with data minimization, as the user does not have to spread his login credentials all over the web.

In Figure 2.5 I have shown an example from Mozilla on how to integrate secure information into the browser. Creating this prototype I became aware of the need for such solutions, as nothing would hinder my application to present a fake Facebook login page and phish the user's credentials. It becomes obvious, that the layers of the digital ecosystem I described in Figure 2.2 need to be developed holistically to communicate where secure information comes

from. Thinking of user interfaces as an afterthought will lead to the problem I described here.

Because OAuth is not well understood by users, there is a lot of mistrust towards the system, and rightly so, seeing how much access can be gained. I would argue that this is also an interface problem that will have to be resolved using consistent user interfaces and mechanisms.

5.5.1 *Experimentation*

I developed the shape for the identity visualization through a lot of experimentation with the prototype. I initially sketched variations by hand, but as soon as I had to parse real data and prepare it for the visualization, I noticed several deficiencies. I created a "lab" to experiment with the shape through sliders in order to see what it looks like using differnt data.

A large part of the experimentation went into preparing the data to see what could be expressed and what parts of the data should make up a dimension of the shape.

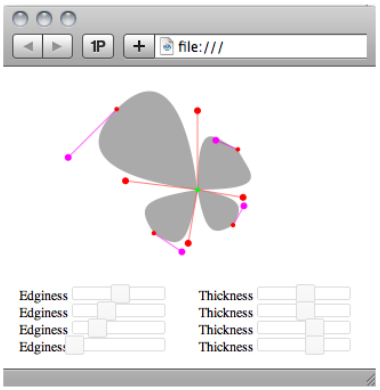

Figure 5.9: Developing the shape was done both by sketching and programming.

Figure 5.10: An early typological analysis of the shape showed several deficiencies like the form falling apart or dimensions with low values vanishing.

5.5.2 *Evaluation*

At this experimental stage of the visualization, it would not have made sense to do a quantitative study of the shape. I also wouldn't have had enough user data to construct a realistic visualization. So I created a website to gather some qualitative data.

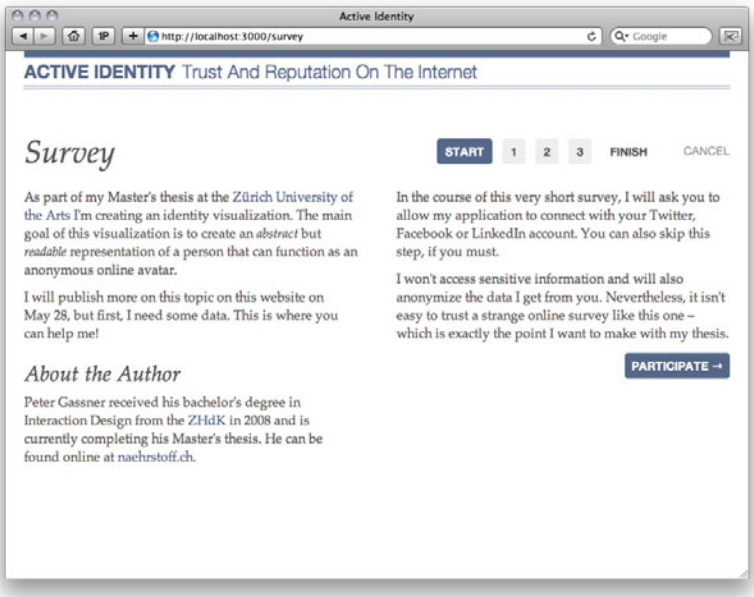

Figure 5.11: The survey had three steps. In the last step, the participants were presented with their personal identity visualization.

An online survey can of course not provide the same information that a face-to-face user study would provide. However, because I required the participants to connect to their Facebook and Twitter profiles, I wanted them to experience this at home. Besides some

multiple choice options on specific questions I had, I added a lot of free-form text areas in order to receive more conversational and open feedback.

I sent this survey to a few selected people because I wanted to know them personally to be able to judge the outcome of the visualization. Also, I didn't want to ask strangers to give me authorization to their Facebook accounts, because this would give me access to a lot of sensitive information.

For this survey, I created six example identities, from which the test subject had to pick three. I tried to make these balanced, so that every identity had good and bad aspects. This is of course "balanced" with regard to my opinion. These figures are depicted in Figure 5.12. Here is my reasoning behind them:

Figure A An identity with a lot of activity, but a bad activity trend, indicating that a lot of activity has been added recently.

Figure B Good reputation, but few sources (the white dot), indicating too little and thus maybe not trustworthy data.

Figure C Lower than average reputation, but many social contacts and activities.

Figure D Everything fine, except for the extremely low activity.

Figure E Too good to be true?

Figure F All parameters are average, the basic shape.

I was interested in whether the main dimensions (the shape) would be understood and how the warnings (few sources, suspicious activity trend, extremely low value) would influence peoples' decisions. Figure C specifically targets the reputation dimension: would people still consider this figure, even though the reputation is low. I wanted to see how people react to this "weighty" term. The results are shown in Figure 5.12, with each participant's choices and remarks listed separately.

Looking at these results, even though they are not statistically relevant, we can see the following:

- People understood that figure E is a figure with good values. Many who chose E also chose figure F, which shows an average identity and thus should be ok. Balance and symmetry were stated as reasons, something I hoped for when designing the shape: if a shape deviates from the basic shape, it becomes asymmetrical and thus less trustworthy.
- People did not regard the warning about few sources in figure B. They either did not understand it, or valued reputation and the coloring more (some stated their bias towards reputation).
- I was surprised that figure A was chosen, because it is colored red. Again, balanced was a criteria and "active contributors to

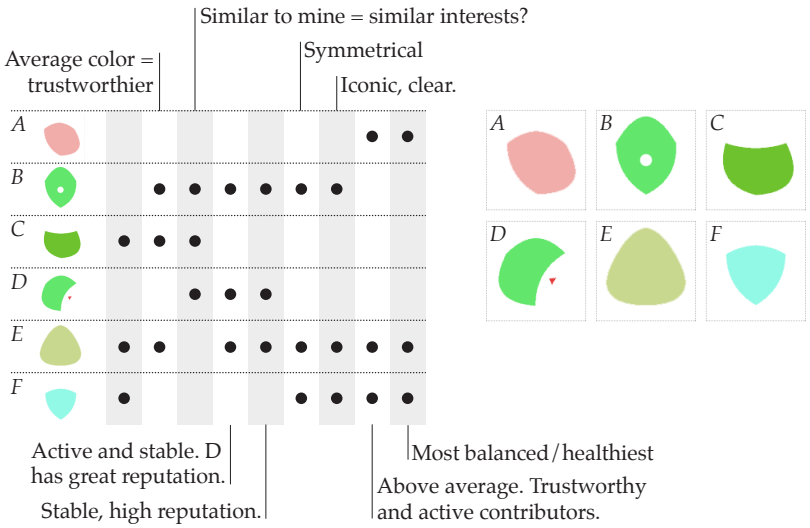

Figure 5.12: The participants were asked to choose three out of these six figures. The parameters of the figure were explained to them before. The results of the nine participants are shown individually including their comments on their choices.

communities". This latter argument is indeed correct, but should be taken with a grain of salt, because the coloring indicates a drastic activity increase recently, leading to this good value.

Half of the participants stated their wish for seeing a portrait photograph of the abstracted person, even though they were aware that it could be fake. One person stated that he was not content with the outcome of their visualization and would like to have more control over it.

In the end, I created a mini-topology (Figure 5.13) of the participants, based only on the data they provided. Even though the data set was very low, we can see that groups can be created.

Figure 5.13: A mini-typology of the identity visualizations of the participants.

5.6 *Scenario: Person-To-Person Lending*

To conclude this chapter, I want to come back to person-to-person lending again, to see how the created visualization could be applied to it. These scenarios would have to be tried out to see how people accept them.

Figure 5.14 shows the basic information a debtor has to make available to a creditor. Using the Core Services provided by the digital ecosystem, this information could be supplied from various sources digitally.

Personal statements like the description at the bottom or the portrait photograph could be validated by looking at the Active Identity visualization, which can give hints about a user's reputation and social activity.

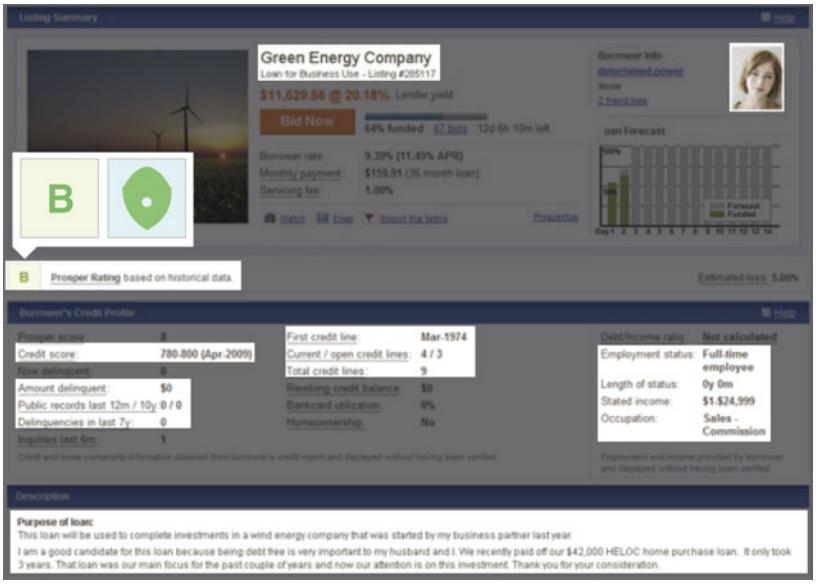

Figure 5.14: The woman in the top right requests a credit to fund a Green Energy Company. (Prosper.com)

Figure 5.15 shows a typical list of credit requests, except that it has been shortened to three items only. To give a creditor a quick overview of the trustworthiness of the debtor, an Active Identity visualization could be supplied alongside the listings.

Figure 5.15: A list of credit requests. (Prosper.com)

Instead of having tabular listings, the lending landscape could be visualized as a whole as shown in Figure 5.16. Through zooming and panning – similar to a map application – creditors and debtors could browse the loans, finding other people (anonymously). Heer and Boyd have shown how social networks could be visualized as a whole, including different views to show sub-communities.[6] But instead of using Gravatars to depict people, the Active Identity visualization could provide representations that are better scannable from a distance and also tell something about the person behind it.

[6] Jeffrey Heer and danah boyd. Vizster: Visualizing Online Social Networks. Sep 2005. URL `http://hci.stanford.edu/jheer/projects/vizster/`

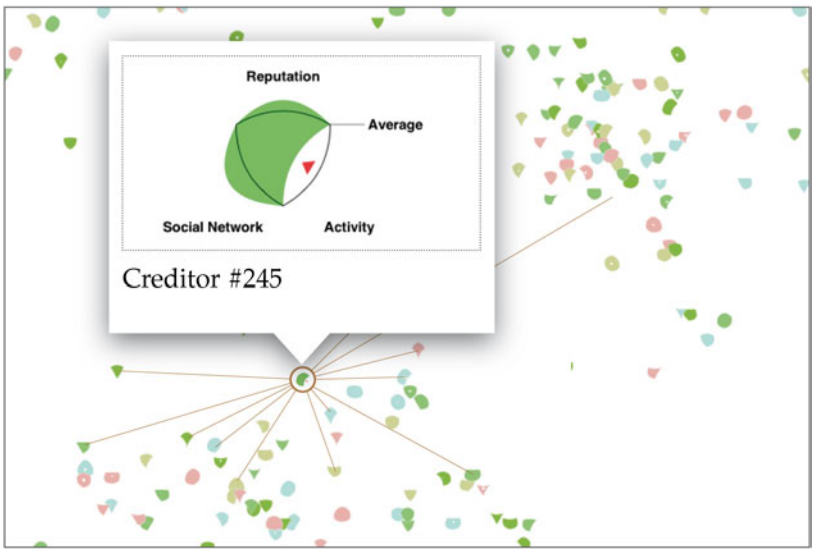

Figure 5.16: Mock-up showing a creditor and the debtors he has invested money in.

As I have shown in Chapter 5.4, typography and spatial positioning could be specially adapted for the use case of person-to-person lending, as there may be parameters that are relevant to this platform only. The pop-overs could display information about a specific debtor, the loan he seeks and maybe previous loans. Also, there should be ways to filter through this large amount of data, not only by grouping, but maybe also with more specific controls to show or hide certain data points.

6
Discussion

I have shown a top-down approach to creating a secure digital ecosystem. A main criterion for such an ecosystem is openness. I described how openness promotes evolutionary processes that can benefit everyone. I see a need for such a system because the Internet is more and more becoming an integral part of our society. Yet today, it is difficult to trust other people on the Internet because their Identity cannot be verified.

Using person-to-person lending platforms as an example, I have shown that we can think differently about money in such a global and ubiquitous medium as the Internet. The Internet allows new marketplaces to be created with ease. There, people with similar interests can meet to make exchanges.

However, making money transactions with strangers requires high levels of trust in technology as well as people. I have shown how a digital ecosystem can provide this trust by clearly separating concerns and putting the user in control. Every secure transaction has to be initiated by the user and is, through use of cryptography, tamperproof. Protecting the privacy of the user is a main goal and has been proposed to be achieved through anonymous and derived claims, that only contain the minimally necessary information.

To communicate these mechanisms to the user, I have created an identity visualization that is based on the principles of digital ecosystems. The visualization is called *Active Identity* to clarify that it is not automatically aggregated but has to be actively curated by the user. The visualization is fact based, the user can not make up the data. However, he can control whether a piece of information enters the visualization or not. This visualization also encourages the user to think differently about his online identity.

Bibliography

Barbara Barkhausen. Tauschgeschäfte. *brand eins 03/07*, Feb 2007.

N Berggren, H Jordahl, and P Poutvaara. The Looks of a Winner: Beauty and Electoral Success. *Journal of Public Economics*, Jan 2009.

Fabian Blaesi. Banking 2.0: Strukturelle Eigenschaften und Vertrauensbildung beim Social Lending. Dec 2008.

Stefan A Brands. *Rethinking Public Key Infrastructures and Digital Certificates*. 2000.

Gerard Briscoe and Philippe De Wilde. Digital Ecosystems: Evolving Service-Oriented Architectures. Oct 2009.

K Cameron, R Posch, and Kai Rannenberg. Proposal for a Common Identity Framework: A User-Centric Identity metasystem. *The Future of Identity in the Information Society*, May 2008.

David Cox, Thomas L Kilgore, Tiffany Purdy, and Rekha Sampath. Catalysts for Change: The Implications of Gen-Y Consumers for Banks. *Deloitte*, Apr 2008.

danah boyd. Faceted Id/entity. Aug 2001.

danah boyd. Making Sense of Privacy and Publicity. *SXSW, Austin, Texas*, Mar 2010. URL `http://www.danah.org/papers/talks/2010/SXSW2010.html`.

R Dhamija, JD Tygar, and M Hearst. Why Phishing Works. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, page 590, 2006.

P DiGioia and Paul Dourish. Social Navigation as a Model for Usable Security. *Proceedings of the 2005 Symposium on Usable Privacy and Security*, 2005.

Jyri Engeström. Why Some Social Network Services Work and Others Don't. Apr 2005. URL `http://www.zengestrom.com/blog/2005/04/why-some-social.html`.

J Goecks and Elizabeth Mynatt. Using Social Methods to Support Privacy Management. *Security & Usability: Designing Secure Systems That People Can Use*, 2005.

Erving Goffman. *The Presentation of Self in Everyday Life*. 1959.

Marit Hansen. Me, Myself and I! Manage your Identities Safely. Dec 2007.

Jeffrey Heer and danah boyd. Vizster: Visualizing Online Social Networks. Sep 2005. URL http://hci.stanford.edu/jheer/projects/vizster/.

J D Lasica. Identity in the Age of Cloud Computing. *The Aspen Institute, Washington, DC*, Apr 2009.

R Leenes, J Schallaböck, and M Hansen. Prime White Paper V2. *PRIME Project*, 2007. URL http://www.prime-project.eu/.

Konrad Paul Liessmann. Eine Kleine Philosophie des Geldes. *Einführungsvortrag, Lech am Arlberg*, Sep 2008. URL http://www.philosophicum.com.

Bernard A Lietaer. Das Geld der Zukunft. *Riemann Verlag, München*, 2002.

Niklas Luhmann. Vertrauen: ein Mechanismus der Reduktion Sozialer Komplexität. *Lucius und Lucius, Stuttgart*, (4), May 1968.

Scott McCloud. Understanding Comics: The Invisible Art. *Harper Paperbacks, New York*, Apr 1993.

Francesco Nachira. Towards a Network of Digital Business Ecosystems Fostering the Local Development. May 2002.

Harold Noonan. Identity. *Stanford Encyclopedia of Philosophy*, Nov 2009. URL http://plato.stanford.edu/entries/identity/.

Robert Ezra Park. Race and Culture. *Glencoe, Ill.: The Free Press*, 1950.

Aza Raskin. Identity in the Browser (firefox). *Aza's Thoughts*, Nov 2009. URL http://www.azarask.in/blog/post/identity-in-the-browser-firefox/.

Jürgen Schmidhuber. Facial Beauty and Fractal Geometry. Jan 1998.

The Eclipse Foundation. Higgins Overview. May 2009.

Frederic Vester. Der Wert eines Vogels. *Kösel, München*, 1987.

Colin Ware. Information Visualization: Perception for Design. *Morgan Kaufmann*, 2000.

weBank. Peer-to-Peer Finance Report. Mar 2009. URL http://webank.org.uk/?page_id=73.

Collette Wright and Michael K Hulme. Internet Based Social Lending: Past, Present and Future. Nov 2006.