



INTERNET LIFE VERIFICATION

Using social data as part of the identify verification process.

Internet Life Verification White Paper

THE OPEN IDENTITY EXCHANGE | BLUE MARBLE RESEARCH

By Nicola Turnill and Emma Partridge, Blue Marble Research

January 2014

Contributors

The following companies have contributed to the findings presented within the paper:



Executive Summary

This paper summarises the findings from an investigative project on the subject of internet life verification (ILV). ILV is the process by which an individual's identity can be verified through analysis of their online activity including social media use. By using the predicative algorithms in these systems it can be determined whether a social media account is likely to relate to a 'real' person. ILV provides a potentially powerful new pathway to complement traditional identity verification methods such as document validation and biographical electronic footprint checking. The project investigated how citizens might use ILV when establishing a trustworthy digital identity for access to online services. This is likely to be of particular benefit to people who do not have some of the traditional means of identity evidence available to them ('thin file' individuals such as recent immigrants and young people - as opposed to those who have 'thick files' with longer, deeper identity history and evidence with which to validate against). This paper evaluates the use of ILV through social media (in this case restricted to Facebook) to provide evidence of a citizen's on-going existence within identity proofing processes and considers likelihood of use, usability, identity verification and commercial feasibility.

It is concluded that ILV using social media login does have a utility within identity processes. The use of ILV is commercially viable with a lower price point than alternative means of identity proofing and that the process has the ability to distinguish between "real" and "fake" social media accounts. This could provide initial low cost screening prior to going out to more costly identity attribute verification.

However, it was clear consumers are unfamiliar with the ILV process within Government transactions and there is evidence of a lack of understanding. In the context of these types of transactions test subjects expressed concern about using a Facebook login, and privacy issues were raised. These concerns do not mean that social media cannot be used within identity verification processes but there is a need for further education and explanation for the user relating to the use, privacy, and an increased understanding about what user data is, or is not being shared. Increasing adoption of ILV in other sectors such eCommerce means there is likely to be a better user understanding in the future.

The use of ILV would benefit from further testing to reach additional conclusions regarding the likelihood of use in certain demographics, particularly amongst those with 'thin files'. Additional testing should also evaluate the use of alternative social media networks (e.g. Twitter and LinkedIn) and define the levels at which the 'confidence score' returned is deemed acceptable in verifying identity.

Table of Contents

1. **What is Internet Life Verification?**
 - Overview
 - Current use of ILV
 - Why use ILV?
 - Future use of ILV
2. **The ILV Alpha Project**
 - Background and hypothesis
 - Objectives and output
 - Methodology and sample
3. **Findings**
 - ILV and identity assurance
 - Likelihood to log-in using Facebook
 - Usability and customer experience
4. **Using ILV on gov.uk**
5. **Commercial considerations**
6. **Conclusions**

1. What is Internet Life Verification?

Overview

ILV is the process by which an individual's identity can be verified through analysis of their online activity (e.g. social network interactions, crowd sourcing, behavioural data and internet repositories). 73% of UK citizens now have internet access¹ and 50% use social media² including 90% coverage amongst 16-24 year olds³. Social media sites contain extensive sources of information about individuals. Using this information to verify identity has the potential to allow access to a wider range of online services, simplify processes, improve security, prevent fraud and increase inclusivity. This Project examined the potential use of social media (specifically Facebook) in verifying digital identity.

Current use of ILV

Though relatively new to the identity verification field, ILV is increasingly being used as an alternative to traditional paper or face-to-face means of verifying identity. Early adopters include apartment sharing site Airbnb⁴, credit reference agencies such as Equifax and Experian and payment companies like Ebay and PayPal⁵. ILV has the potential to be used to assert creditworthiness and identify fraudsters^{6,7} and for anti-money laundering⁷. Its use is highly likely to increase as the need for identity validation grows.

Why use ILV?

There is no single, authoritative source for validating identity. Even passport issuing organisations are susceptible to fraudsters. Furthermore not all citizens may have a passport or are registered on the electoral roll. The widespread use of social media provides another tool for identity

¹ ONS. Statistical Bulletin. Internet Access - Households and Individuals, 2013 http://www.ons.gov.uk/ons/dcp171778_322713.pdf

² EMarketers. Social networking to reach half the UK population this year. July 2013. www.emarketer.com/Article/Social-Networking-Reach-Half-UK-Population-This-Year/1010032#SuT4opexMmkEoSd.99

³ ONS. Social networking. The UK as a leader in Europe, June 2013.

⁴ Hempel, J. Airbnb to launch identity-verification tool. CNN. April 2013. <http://tech.fortune.cnn.com/2013/04/30/airbnb-to-launch-identity-verification-tool/>

⁵ Kucera, D. Facebook posts help credit bureaus sniff out fraudsters. Bloomberg, May 2013. www.bloomberg.com/news/2013-05-30/facebook-posts-help-credit-bureaus-sniff-out-fraudsters.html

⁶ Zoot. #Trending: Social media data (Is the future now?). April 2013. www.zootweb.com/blog/index.php/trending-social-media-data-future/

⁷ Pitter, A. Social media data: The next frontier for identity verification? Global Data Company. April 2013 www.globaldatacompany.wordpress.com/2013/04/04/social-media-data-the-next-frontier-for-identity-verification/

'Thin' and 'thick' file individuals

The majority of UK citizens have **'thick' files** which can be used to verify their identity e.g. financial records, council tax or utility bills in their name.

In contrast, **'thin' file** individuals do not have some of the traditional forms of identity credentials at their disposal. Examples include young people, ex-military and recent immigrants. These groups may be more likely to benefit from using social media as a means of identity verification.

GPG45 - Activity Evidence Categories

- **Citizen (C)**. Evidence which demonstrates an interaction between an individual and a Public Authority as a citizen of the state e.g. electoral roll activity.
- **Money (M)**. Evidence that demonstrates the individual's financial life e.g. credit card payments.
- **Living (L)**. Evidence that demonstrates where the individual lives, their working life and what they consume e.g. use of social media, land registry entry, national pupil database entry.

verification in the form of ILV It will particularly benefit 'thin file' individuals (left) who may have experienced difficulty entering the digital world and who might otherwise be identity proofed to lower levels. Ultimately ILV increases the potential for every UK citizen to have access to online services.

Future use of ILV

In 2013 the public and private sector collaborated to create a set of standards called the Good Practice Guide (GPG), which could be adopted by any organisation looking to enter the digital world. The GPG45⁸ deals specifically with the identity proofing process. Within these guides ILV currently sits within the 'Activity Evidence' component of the process. This verifies the presence of historical and/or continued activity, thereby providing an 'Assured Identity'. Activity Evidence must be drawn from 'evidence categories' - citizen, money and living (CML; see left) and it is thought that ILV sits within the 'living' category.

If in the future citizens chose to leverage social media to help establish their identity, the process uses algorithms to analyse data trends and characteristics (e.g. networks, frequency of use) in order to distinguish "real" from 'fake' profiles and assign an ILV user 'confidence score'. The system does not require individual level data (e.g. wall posts, 'likes', photo tags), rather it provides conclusions about identity based on interactions across time, using multiple sources and correlations. A further benefit is that accuracy will increase over time as more people are verified, because confidence scores are bolstered by connections to other verified individuals.

⁸ Cabinet Office CESG. Good Practice Guide no. 45. Requirements for secure delivery of online public services. Issue 1.1. 2012.

About Projects

In order to hasten adoption of the IDAP Framework there is a practical and strategic opportunity to leverage OIX domain expertise. OIX facilitates and coordinate the rapid formation and deployment of *White Papers, Discovery* and *Alpha Projects* in an agile manner.

These are defined as small scale, low risk assessments, analysis or tests of interoperable components that address the key challenges of the IDAP goal to create convenient, secure, and privacy-enhancing digital transactions.

Objectives and Output

- **Effectiveness.** Assess the effectiveness in relation to identity assurance and overall user experience.
- **Cost.** Evidence of the relative costs of ILV.
- **Commercial.** Assess the commercial benefits.

2. The ILV Project

Background and hypothesis

This project was initiated on behalf of the Identity Assurance Programme (IDAP) initiative between the UK Cabinet Office and the Open Identity Exchange (OIX). The project supports IDAP goals, which in turn support the “digital by default” strategy for citizen-government transactions. The following hypothesis is tested: *“How effective and cost efficient will output from the ‘internet life’ category be in verifying identity?”* Effectiveness can be viewed from user, IDP and vendor perspectives and includes meeting assurance standards, providing a positive user experience and cost efficiency.

Objectives and output

The key objectives of this Project are to:

- i) Assess effectiveness of ILV, focusing on what kind of identity verification is achieved, the overall user experience and the person’s propensity to use ILV as evidence of identity.
- ii) Provide a high-level assessment of the relative costs of ILV.
- iii) Assess the commercial benefits.

This paper presents key findings from consumer testing^{9,10} and commercial analysis¹¹ and provides independent conclusions.

Method Summary

ILV was tested across a number of user journeys in the context of a government digital transaction. This tested if citizens had a propensity to use ILV, namely Facebook as part of the identity proofing process, the usability aspects of this and additionally the affect on identity proofing and assurance.

⁹ Verizon. Internet Life verification OIX Alpha Project report. August 2013.

¹⁰ Brewer, D. Mydex OIX Alpha – ILV. September 2013

¹¹ Lindley, E. Innovate Identity. ILV commercial report. August 2013

Figure 1: Scores returned from journey 2 testing

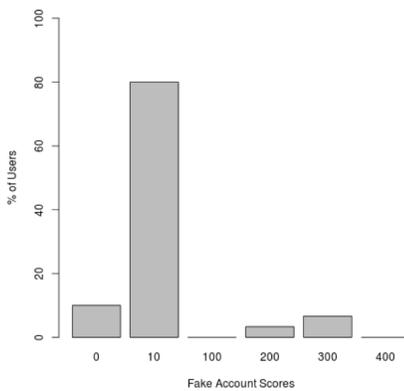
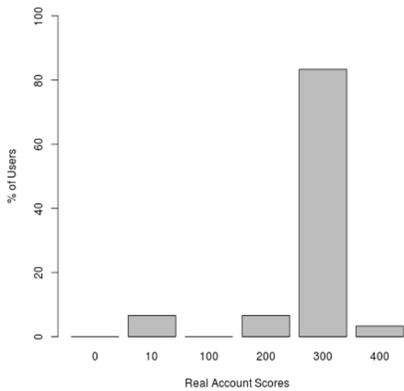


Table 1: Methodology and sampling

Test No	Method	Sample	Scenario type	Recruitment Method	Facebook spec
1	Moderated	10	Renew driving license	Undefined	Facebook user
2	Online virtual lab	30	Book driving license test	Specialist agency Useresting.com	Main & 'atypical' Facebook account
3	Moderated	5	Universal credit claim	Thames Valley Housing Facebook page	Use of a Facebook 'persona'

3. Findings

ILV and identity assurance

The key finding from an identity assurance perspective was to consider the viability of the confidence scores from ILV returned in evaluating identity.

During one of the user journeys tested, participants were asked to try and login with their genuine personal Facebook account and also to login with a fake Facebook account they had created.

It can be seen from the graphs (left) that in 92% of cases ILV was able to score these accounts differently. Below are the specified descriptions from the ILV supplier about what constitutes a “real” or “fake” Facebook account.

1. Authentic Account “real” is defined as an account, which demonstrates numerous (variable) characteristics known to be associated with legitimate social network profiles, and it very difficult to duplicate. This account profile correlate to a legitimate offline identity, and the bearer has been authenticated as its owner.
2. Inauthentic Account “fake” is defined as an account with numerous characteristics know to be associated with illegitimate social network profiles, and can be easily duplicated. This account may not correlate to

Figure 2: The 'Login with Facebook' button that can be seen on many websites

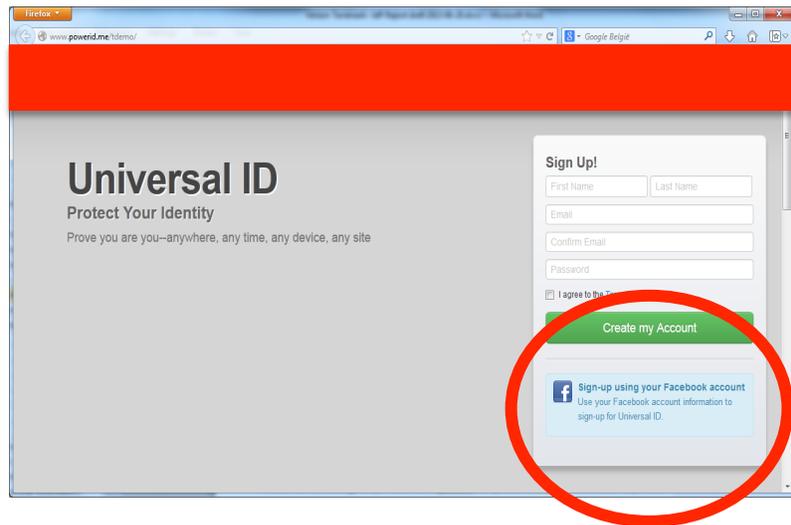


an offline identity, or may correlate to an offline identity associated with fraudulent activity. Further verification is required.

It was established that ILV had a high degree of accuracy in distinguishing through the predictive algorithms the genuine or “real” accounts and the “fake” ones provided by the test participants.

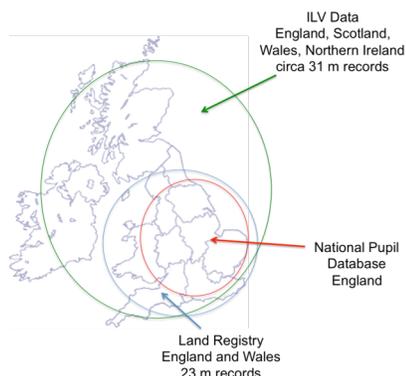
Likelihood to login using Facebook

Login with Facebook can be described as a personalised way for people to sign in to websites and applications using their existing social login information. This functionality is aimed at speeding up the registration process for citizens because they do not have to fill out all the usual details, such as name and address. For organisations it enables the build of a customer registration and login system quickly.



It was established that within the context of these public service transaction types there was distrust of the use of Facebook, concerns about use of personal data and evidence of a lack of understanding. Many people did not notice the option to sign in with Facebook. Participants also spent little time reading instructions, which may have impacted understanding. The concept of ILV within an identity process is unfamiliar to some consumers resulting in distrust.

Figure 4: Data coverage within the 'living' (L) category of Activity Evidence



Using ILV on gov.uk

The reluctance to use social media to verify identity was due to a misunderstanding of how ILV would work in this context. To increase likelihood of use there needs to be further education and explanation relating to the use of ILV. This should include:

- Highlighting the benefits in terms of ease and convenience to the citizen.
- Clear messaging in relation to privacy, data protection and the use of personal information.
- Explanation of how ILV works, through the analysis of data trends and correlations.

4. Commercial considerations

The use of ILV means that an increased number of UK citizens will be able to access government digital services. As outlined in the Martha Lane Fox Report¹², increasing the number of people online ultimately increases the potential for cost savings. Table 2 shows commercial comparisons of the data sources within the Activity Evidence component of the GPG45. A combination of these would be required within the identity proofing process, and as shown, there are clear cost benefits to using ILV methods over the alternatives. Cost savings could therefore result from ensuring that ILV methods are utilised in the first instance, before offering more costly mechanisms

Table 2: Commercial comparisons: 'living' data within Activity Evidence

Category	Data Source	Indicative Price Point Per Transaction (Low – High Volume)
C	Electoral Roll	£0.05 - £0.25*
M	Credit Data for Identity Purposes	£0.15 - £1.00*
L	Land Registry	£0.19 - £0.28**
	National Pupil Database	Free***
	Internet LifeVerification	Free - £0.13

* Anonymised ranged pricing taken from multiple vendors and aggregators of available data
 **Commercial prices i.e. not under Government Data Sharing Agreement – these prices would be reduced if the IDP's come under the Government Data Sharing Agreement
 *** Access only if approval granted under the Government Data Sharing Agreement. This data is not available outside of this agreement.

¹² Martha Lane Fox Report Directgov 2010 and Beyond: Revolution not Evolution (2010)

Conclusions

- **Viabile.** In the context of identity in it's ability to spot real and fake users
- **Cost effective.** In comparison to other sources
- **Fewer passwords.** Could be required in the future with this method
- **Education.** Will be important in adoption
- **Privacy.** Was critical for users
- **Transaction context.** Is going to be important to uptake

Different data sources have different prices; they also have different coverage of the UK population and differing ease to obtain fraudulently.

GPGs look for a spread of evidence of a person's identity across different categories of data. This makes it harder for a fraudster who might have successfully attained one piece of identity evidence e.g. a passport in a false identity but has not attempted to claim benefits (or vice versa). The 'living' category is a new and potentially rich category of evidence. We need to understand the systemic ways in which it can be used and abused and how much weight it should hold relative to other categories. Those making commercial decisions will wish to understand how cost effective it is.

Conclusion

There is no single answer to proving identity in a digital transaction, even traditional methods are susceptible to fraud. The new world of "big data" allows access to more information, which could be used, for identity, such as ILV. This project makes clear that the context of the transaction is important to the use of ILV; and that within this test of Government transactions, ILV was not an attractive method to the user due to the lack of understanding.

Increasing adoption of ILV in other sectors could mean an increased understanding of ILV in the future. It is therefore concluded that once further refined, ILV does have a utility within identity verification processes. ILV has been shown to be commercially viable when considering price point in relation to traditional means of identity verification. The ILV process is effective in distinguishing between "real" and "fake" social media accounts and can benefit consumers in terms of convenience, ease and the need to remember fewer passwords.

Additionally, ILV provides an alternative verification pathway due to the coverage of the data itself; this is likely to be particularly valuable to

Conclusions

- **Further testing.** Larger sample sizes
- **Other social media testing.** Twitter, LinkedIn and others
- **Specific demographics.** Should be targeted
- **Inclusion into identity evidence category.** Would widen the scope

those with ‘thin’ files who may have previously experienced difficulties accessing digital services. ILV therefore has the potential to improve inclusivity.

In evaluating the data collected as part of this Alpha Project it is important to acknowledge that this research is qualitative in nature and drawn from small samples that are neither representative of the population as a whole or the key target audience for ILV. The findings are therefore directional rather than absolute and would benefit from additional quantitative testing with larger samples once the user proposition issues have been addressed. This would allow further refinement and more valid data.

This testing should evaluate the likelihood of use of alternative social media accounts (e.g. Twitter and LinkedIn) and should include a sample of ‘thin’ file individuals. Further testing is also needed to ascertain the viability of the confidence scores returned when distinguishing between “fake” and “real” accounts and to define acceptability boundaries.

Further research should be completed to ascertain if the use of ILV could be widened in the context of GPG45. Assuming successful inclusion within Activity Evidence, in the future ILV might also be utilised within ‘Identity Evidence’ area of GPG45. It would thus be possible to make further commercial comparisons between ILV and comparable data sources within this additional category.

Glossary | Terms

CML: Citizen, Money, Living (components of the GPG45 ‘Activity Evidence’)

GDS: Government Digital Service

GPG: Good practice guide

IDAP: Identity Assurance Programme

IDP: Identity provider (e.g. Mydex and Verizon)

IDSG: Identity Assurance Programme Identity Steering Group

ILV: Internet Life verification

OIX: Open Identity Exchange organization