Driving innovation together

Remote Vetting PoC – the design

for SURFsecureID

Authors:	Laura Claas, Bob Hulsebosch, Maarten Wegdam
Reviewers:	Innovalor: Ines Duits, Willem Noort; SURFnet: Peter Clijsters, Joost van Dijk, Pieter van der Meulen
Version:	1.0
Date:	13-12-2019

Moreelsepark 48
3511 EP Utrecht
The Netherlands

PO Box 19035
3501 DA Utrecht
The Netherlands

+31 (0) 88 - 787 30 00
info@surf.nl
www.surf.nl

Synopsis

The report describes the design for a Proof-of-Concept with remote vetting for SURFsecureID. Specifically, it describes how iDIN, ReadID (NFC passport) and IRMA (DigiD/BRP) can be used to make it possible to remotely identify a user that already has a SURFconext identity. There is specific attention for the matching challenges related to using iDIN, ReadID or IRMA, meaning how to match the personal attributes from the existing SURFconext identity with those from iDIN, ReadID or IRMA.

Table of contents

Management summary

Background

SURFsecureID allows SURFconext users to obtain a second factor authentication token, providing additional identity and authentication assurance on top of the institutional username and password-based account. Getting a valid token consists of two main processes, namely 1) a self-service registration process that allows the user to select a token and to link it to their institutional account; and 2) a face-to-face identity vetting process at the registration desk of user's institution to activate the token. This face-to-face identity vetting process is used to get the required identity assurance, but it is highly impractical for remote users, since they do not always reside in the vicinity of the physical registration desk. In addition, it does not scale well for large groups of users.

Therefore, SURFnet is looking for remote identity vetting solutions for SURFsecureID. Remote vetting is the remote, or location independent, identity vetting process in which the identity of a user is bound to the second factor authentication means and to the institutional account. This vetting process can, in principle, be performed either online or in a face-to-face context not bound to the institution's location. In previous research InnoValor analysed several possibilities for remote vetting[1]. In this report, three remote vetting solutions, namely iDIN, ReadID and IRMA, are investigated further, laying the groundwork for a Proof-of-Concept (PoC) and/or pilot with these solutions.

iDIN is provided by the Dutch banks, to authenticate using the online banking credentials. InnoValor's ReadID is a mobile solution leveraging the NFC capability of smartphones to remotely read and verify the RFID chip in modern identity documents. IRMA is a mobile-based decentral and privacy-friendly authentication solution that leverages, amongst others, the Dutch government DigiD authentication solution in combination with the Dutch *Basisregistratie Personen* (BRP, national register of all inhabitants).

Goal

The goal of this report is to further investigate and design the details of remote vetting for SURFsecureID with iDIN, IRMA and ReadID. This includes the following sub-goals:

- To analyse whether IRMA is suitable as a means for remote vetting;
- To analyse the quality of the matching of identity data from the institutions with that from iDIN, ReadID and IRMA;
- To design a registration process with remote vetting by means of iDIN, ReadID and IRMA.

This report is input for a PoC which SURFnet will conduct, to implement and evaluate the above three remote vetting solutions. This PoC is out of scope for this report.

The research underlying this report was mainly done in the first halve of 2019, and reflects the status of around August 2019.

[1] The research report is available at https://www.surf.nl/files/2019-02/report%20remote%20vetting%20for%20surfconext%20strong%20authentication.pdf

Approach

The research is divided in three distinct activities: IRMA analysis, matching analysis and functional design. Each activity's approach and results are described below. Overall, a combination of desk research, expert interviews, small-scale experimentation, functional design, and UI design was used.

Besides doing the actual research described in this report, InnoValor is also the vendor for the ReadID identity verification software. In the actual PoC the role of InnoValor will be minimal and especially the subsequent evaluation of the PoC will be done by SURFnet and not InnoValor.

IRMA analysis

IRMA was not part of the previous research for suitable remote vetting methods, since it did not exist in a form suitable for remote vetting at that point in time. This is now offered by the Privacy by Design foundation. To decide if it should be added to the PoC, next to iDIN and ReadID, IRMA was analysed against the nine criteria formulated in the previous research on remote vetting possibilities.

This analysis showed that IRMA is an interesting remote vetting solution, especially because of the unique option to leverage BRP attributes that it offers via the Gemeente Nijmegen using DigiD. But it has several uncertainties. First, since IRMA relays and caches the attributes from other trusted sources via a decentral architecture, the trust in the attributes is somewhat reduced depending how long ago the attributes were cached. However, since most attributes are relatively static in time this seems not a big issue for SURFsecureID. Second, IRMA's business model is currently unclear; can IRMA continue to ensure that the attributes are provided free of charge in the future? Lastly, if it is possible for the IRMA app to get BRP attributes from the Gemeente Nijmegen, beyond the current pilot, then by extension SURFsecureID might be able to do so without involvement of the Privacy by Design foundation as well. This is under the assumption that municipalities are willing to cooperate with other solution providers in this area.

It was therefore decided to include IRMA in the PoC. This provides hands-on experience with IRMA, especially in the DigiD/BRP possibility it offers, and allows for evaluation of the quality of the attributes for the remote vetting process.

Identity Matching analysis

Users are registered with their first and last name by their institutes. SURFsecureID receives these first and last name attributes though SURFconext from the identity provider (IDP) of each institution[2]. All three remote vetting solutions also provide these same attributes. But these may not match exactly; e.g., shortened names, incorrect registrations at (especially the identity provider of) the institutions or iDIN (e.g., name of partner contrary to actual name), diacritics, missing first names (only initials). In addition, first and surname are not unique, so additional attributes may be needed. A matching analysis was therefore conducted to assess the quality of matching between SURFconext/institutions and the three remote vetting solutions. For this, desk research was combined with a small-scale experiment in which participants requested attributes from iDIN, legal identity documents via ReadID, the BRP via IRMA, and SURFconext IDPs.

The analysis showed that matching identity provider (IDP) identity assertions with assertions provided by iDIN, ReadID or IRMA is not trivial for several reasons. First, the analysis confirmed that IDPs do not provide sufficient attributes to be able to uniquely match identity assertions with each other. Currently only surname

[2] For more information on identity attributes in SURFconext, see
https://wiki.surfnet.nl/display/surfconextdev/Attributes+in+SURFconext

and first name or initials can be used; it is recommended that IDPs also share 'date of birth' as a mandatory attribute and optionally 'gender' to reduce the risk of incorrectly matching someone. Second, there is little homogeneity across the values of the attributes provided, which requires an extensive translation rule set for matching attributes. Since quite some personal information is being (automatically) processed by SURFnet during this matching process it is recommended to conduct a privacy impact assessment on SURFsecureID.

Our recommendation is to do the PoC with a simple rule set for automatic matching, supported by manual quality checks of matching queries by a Registration Authority (RA) when necessary (as is now done for the face-to-face process). This in contrast to either a hard rejection by the automatic matching process, or upfront big investments to reduce the matching fails. If larger amounts of users need to be vetted, a more extensive rule set may be needed and/or a richer set of attributes. This also depends on SURFsecureID's risk appetite concerning the false acceptance rate of the identity matching and on the outcome of the PoC.

Functional design

A functional design was made on how the remote vetting process can be integrated in the existing vetting process, for each of the chosen remote vetting solutions. Several high-level design decisions were formulated, on the basis of which detailed user flows were established. Some key design decisions are:

- The remote vetting process should stay as close as possible to the existing physical identity vetting process.
- The current email verification/activation is excluded from the remote vetting process.
- The user can choose between the three remote vetting options.
- The (automatic) identity matching will not be allowed to stop the process; if the (automatic) identity matching fails, the user can try with a different remote vetting method. If this fails as well, the vetting falls back to an RA. The RA gets digital insight into the vetting process, i.e., the SURFconext and iDIN/ReadID/IRMA attributes, and then makes a decision: approve the matching, block the user, or initiate a retry.

This results in the following overall process:

1. The user logs in at the SURFsecureID self-service registration portal with his federated institutional account.
2. The user selects a second authentication factor (SMS, Tiqr, YubiKey, …) to register, and performs an authentication to prove he owns the token, and to link the token to the institutional account.
3. The user selects one of the remote identification options.
4. The user executes the identification steps.
    a. For iDIN:
        i. the user selects the bank to be used for the iDIN identification and authenticates using iDIN.
        ii. The identity attributes from iDIN to be shared are shown. The user provides his consent to share these attributes with SURFsecureID.
        iii. The user is redirected back to the SURFsecureID self-service portal.
        iv. Meanwhile, in the backend, the attributes are communicated with SURFsecureID.
    b. For ReadID
        i. The user receives an instruction to download and install the ReadID Ready application.
        ii. The app is linked to the user's SURFconext account, either via scanning a QR code shown in the SURFsecureID portal or by clicking an activation link (the latter if the user is using a mobile device and the ReadID Ready app is installed on this same device).
        iii. The user reads his identity document with the ReadID Ready app, via NFC.
        iv. The user performs a selfie-check to confirm he is the rightful holder of the identity document (this step could be skipped but we do not recommend this).

v. The outcomes of the identity document verification and selfie-check are communicated to SURFsecureID.

c. For IRMA:
   i. The user receives an instruction to download or open the IRMA app.
   ii. The user downloads the app and creates an IRMA account.
   iii. The user obtains the identity attributes from the BRP. This is always done to increase security and trust level, i.e., do not used cached attributes that are e.g. 60 days old. This is done as follows:
      1. The user is directed to the website of Gemeente Nijmegen, where he logs in with DigiD SMS or app.
      2. The user consents to importing the BRP-attributes into IRMA.
      3. The user is directed back to IRMA.
   iv. The user shares the BRP identity attributes with SURFsecureID, by scanning a QRcode displayed in the portal.

5. In the backend, the identity (attributes) provided from the chosen identification solution are matched with the SURFconext identity by SURFsecureID. When a match is established, the token is linked to the SURFsecureID (i.e. the user's SURFconext) account.
6. If the matching fails, the process is handed over to the RA who decides.

From the functional design, UI mock-ups were made. For each of the screens it is indicated whether they are to be built from scratch, adapted from the existing process, or from external parties.

Trust levels

A risk analysis was made of several potential risks to the level of assurance of the remote vetting process. For each of the identified risks, mitigating controls have been recommended. Assuming those mitigating controls are in place, we come to the following analysis of levels of assurance for the second factor obtained by the remote vetting processes:

- Remote vetting with iDIN is roughly substantial or SURFnet's Level of Assurance (LoA) 3.
- Remote vetting with ReadID is roughly substantial or LoA 3. The ReadID process most closely resembles the current (non-remote) vetting process.
- IRMA has a lower LoA compared to the other two remote vetting solutions and the current process, mainly because of the usage of DigiD Midden (which is eIDAS Low). But this could change, by enforcing the use of DigiD Substantial.
- The current, non-remote, identity vetting process has a few notable characteristics when it comes to trust levels. First, there is no proper check of the authenticity of the identity documents. Since RA's are untrained and have no special equipment or access to a stolen/lost document database, detecting inauthentic documents is hard. Second, since the process is executed by a human being (the RA), it's more sensitive to social engineering, e.g., someone using a copy of an identity document, or not resembling the face image. Lastly, vetting is a manual process done by a trusted person; despite proper training human errors may still occur.

In conclusion, iDIN and ReadID cater for a remote vetting process that allows for the provisioning of a second authentication factor at roughly assurance level Substantial (or LoA3 in SN terminology), IRMA may not. It would be fruitful to consider assigning several levels of assurance, or at least store which method was used as meta-information for a specific SURFsecureID account.

1 Introduction

1.1 background

In a previous research[3] on the theme of remote vetting done by InnoValor for SURFnet, it was researched how users of SURFsecureID[4] can identify themselves remotely to obtain a second factor token that provides additional identity assurance to their institutional username and password-based account. It gives the users access to cloud-based services that are linked to SURFconext and require stronger forms of authentication than provided by their home institute. Users log in with their institution's account and, as an additional step, are then prompted to confirm their identity with the second factor authentication token. Currently, SURFsecureID gives access to cloud services via three different types of authentication tokens: SMS, Tiqr (smartphone app) or YubiKey (USB hardware token).

Getting a valid token consists of two main processes, namely 1) a self-service registration process that allows the user to select a token; and 2) a face-to-face identity vetting process at the registration desk of user's institution to activate the token. Due to the face-to-face process, the identity vetting is mainly suitable for users that work at the institutional buildings. Users that work elsewhere or abroad are required to travel to the registration desk at the institution, which is highly impractical. Moreover, if the number of users that require a strong authentication solution is limited, the costs and effort of setting up and maintaining a registration desk including trained registration authorities, do not weigh against the benefits. The opposite situation however is equally infeasible; if large amounts of users need to be enrolled in short time (i.e. bulk enrolment), a face-to-face process able to sufficiently handle the bulk will require tremendous time and labour resources. A fully automated process, available as self-service for users, would be preferable.

For these three use cases, i.e., remote users, a limited number of users, and bulk enrolment, SURFsecureID is looking for remote identity vetting solutions. In the above-mentioned previous research InnoValor analysed several possibilities for remote vetting. IDIN (from the Dutch banks) and ReadID (InnoValor's NFC-based mobile identity verification software) were selected as most promising remote vetting solutions. After this research was finished, IRMA emerged as a potential solution. IRMA is therefore analysed in the same manner as the other remote vetting possibilities. In this report these three remote vetting solutions – iDIN, ReadID and IRMA - are investigated further, laying the groundwork for a PoC and, possibly, subsequent pilot with these solutions.

1.1.1 Remote vetting

Remote vetting is the remote, or location independent, alternative to the current identity vetting process done by a Registration Authority (RA) at the institution. In the previous research, options were explored that involved face-to-face identification independent of the physical location of the institutional building (e.g. at the user's own door). These however did not live up to the assessment criteria. The three solutions covered in this report (iDIN, ReadID and IRMA) are all fully online solutions.

[3] Bob Hulsebosch, Maarten Wegdam, Remote Vetting for SURFconext Strong Authentication, December 2017, https://www.surf.nl/en/report-remote-vetting-for-surfsecureid.

[4] For more information see https://www.surf.nl/diensten-en-producten/surfconext/wat-is-surfconext/surfconext-sterke-authenticatie/index.html or https://wiki.surfnet.nl/display/surfconextdev/SURFconext+Strong+Authentication.

Remote vetting can be employed for four distinct use cases:

1. for institutions that have such a limited amount of users, that the cost of an own physical registration desk and RA are not worth it;
2. for institutions with remote Dutch users;
3. for institutions with remote foreign users;
4. for bulk enrolment.

1.2 Goal

In the previous research, iDIN and ReadID were identified as the most interesting solutions for remote vetting to be researched further. IRMA was identified as an additional potential remote vetting solution after this research was done; therefore, it will be subjected to the same assessment as the original longlist of potential remote vetting solutions. In this research iDIN, ReadID and IRMA will be further investigated with the goal of applying them in the remote vetting implementation of SURFsecureID, if appropriate. Therefore, the goal of this current research is to further detail remote vetting for SURFsecureID with iDIN, IRMA and ReadID.

This report is input for the actual PoC and the subsequent evaluation. In this PoC the role of InnoValor will be minimal and especially the subsequent evaluation of the PoC will be done by SURFnet and not InnoValor. This also because InnoValor is the vendor for ReadID and we want to avoid an appearance of conflict of interest.

The sub-goals are described below.

1.2.1 IRMA analysis

The goal is to analyse IRMA on its fitness as a means for remote vetting. IRMA is currently employing a pilot in which access to the BRP (basic registry of persons) is possible. This is provided via the municipality of Nijmegen, after logging in with DigiD. This was not yet possible when writing the previous report on the basis of which iDIN and ReadID were chosen. With the availability of BRP data, IRMA has regained new interest among relying parties. Due to this situation, IRMA seems to be an interesting candidate for reinforcing SURFsecureID's vetting process. Firstly, because it is not dependent on NFC, unlike ReadID. NFC currently[5] only works on Android, IRMA works on Android and iOS. Secondly, the data comes from the same source as ReadID, namely the BRP.

IRMA uses cached personal attributes. Potentially, this may impact the reliability of SURFsecureID's vetting process as attributes may be outdated. The risk that attributes have changed, however, is small since most of them are relatively static (i.e. name, data of birth). IRMA will be evaluated along the same 9 criteria as the long-list of the original research report. On the basis thereof, SURFnet will make an informed decision about not, partly, or wholly including IRMA in the PoC.

1.2.2 Identity Matching analysis

To analyse the quality of the matching of data from the institutions with attributes from iDIN, passport chips (ReadID) and, if added to the PoC, IRMA. This includes analysing the consequences for the reliability of the identities compared to the current process.

[5] When finalising this report, on June 3th 2019 Apple released the beta for iOS 13 which allows third-party access to the APIs of the embedded NFC antenna of iPhones. With this it is possible for ReadID to read document chips on both Android and iPhones. iOS 13 is at the time of writing only available as a developer beta. The production version of iOS 13 is expected in September 2019.

1.2.3 Functional design

Functional design of the registration process with remote vetting in SURFsecureID. This includes the development of UI mock-ups of the new registration process.

1.3 Approach

The research underlying this report was mainly done in the first halve of 2019, and reflect the status of around August 2019.

This report describes the outcome of three main activities, corresponding to three above three sub-goals.

1.3.1 IRMA analysis

An analysis of whether and how IRMA can be used in the remote vetting process for SURFsecureID. This analysis will be done against the nine criteria formulated in the previous research project. Depending on the outcome of this analysis, IRMA will be further included in the PoC.

The outcomes of this activity are described in chapter 2 of this report.

1.3.2 Matching analysis

How reliably can the identities yielded by the chosen means (iDIN, ReadID, IRMA) be matched with the identities of the institutions? To determine this, a pre-PoC and pre-pilot analysis will be done, as to steer the PoC implementation. The analysis will be based upon:

- Studying previous research by SURFnet (e.g. eduID) and conversations with DUO about Studielink
- Analysis of which personal attributes are available from iDIN and ReadID / IRMA BRP and what the quality of these attributes is. This will be done by a combination of studying standards, informal discussions with banks or Betaalvereniging, and a few small-scale experiments (e.g. requesting attributes from some of the people involved in this research project via iDIN at different banks).
- Assessment of the existing matching solutions on the market and with other parties such as RVIG and GovUK.

The outcomes can be found in chapter 3 of this report.

1.3.3 Functional design

A functional design will be made on how the remote vetting process can be integrated with the existing vetting process, for each of the chosen vetting means. The functional design will include UI mock-ups in Balsamiq. In the design, design decisions will be made such as:

- To do, or not to do, a selfie check, and what the consequences of this decision are for the level of assurance;
- To make explicit whether the LoA will be substantial or high, and how this relates to the current vetting process.

Attention will be paid to both understandability as well as trust. Both will be evaluated in a later stage, during the PoC and pilot, on the basis of working systems.

For ReadID the ReadID Ready app will be used; a white label ready-to-use app utilizing NFC for scanning chips of legal identity documents and including selfie-check functionality for holder verification. ReadID Ready app is provided by InnoValor; SURFnet will have to customise it for and integrate it in the vetting process.

The result of the functional design is a set of clickable mock-ups that demonstrate the screen flows and interactions. These serve as a guidance for the implementation of the PoC. The results can be found in chapter 4 of this report. The mock-ups can be found in chapter 5.

2 IRMA analysis

2.1 About IRMA[6]

2.1.1 Background

Since the original analysis of remote vetting solutions in 2017 there is a new candidate: IRMA. IRMA is an Idemix[7] based privacy-friendly identity platform. IRMA has been around quite some years, originally smartcard-based but now mobile-based. IRMA originates from the Radboud University, and many of the people active in the foundation are from Radboud University in Nijmegen, including the chairman of the board Prof. Dr. Bart Jacobs.

What changed is that IRMA is now provided as a service by the Privacy by Design foundation (`https://privacybydesign.foundation`), and this foundation enables users to access Dutch government information (BRP, see below).

2.1.2 Architecture

At a high-level, IRMA is an app which users can use to load one or more identity credentials onto. Credentials can consist of several attributes which can then selectively be shared with websites (called verifier). The issuers provide the actual attributes and signs them. The verifier can verify that an attribute was indeed issued by a certain issuer. The Schema Manager, accessible online and via the app, manages which issuers are available. The user can control which attributes are shared with the verifier, and which are not. These attributes can also be derived attributes, e.g., "over 18" contrary to a date of birth. Figure 1 below gives a representation of the basic architecture:

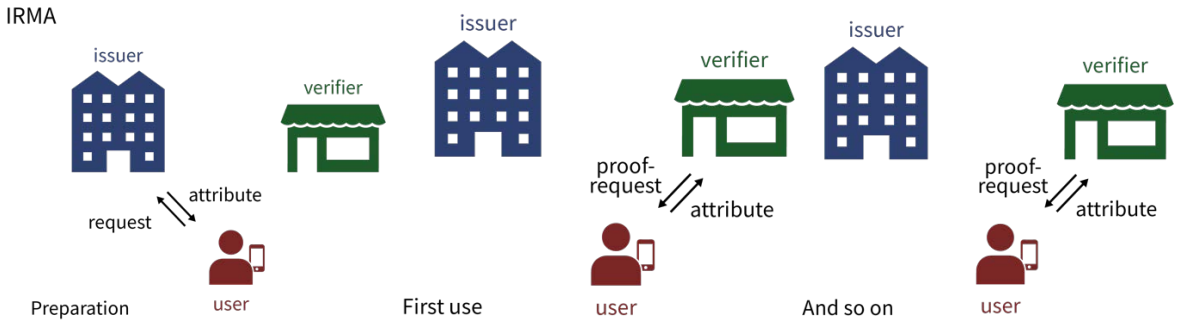


Figure 1: The images show the process of requesting and sharing attributes with different verifiers.[8]

The underlying architecture and protocols work in such a manner that the typical big-brother issue of federated identity systems is prevented: neither the originator of the attributes nor IRMA (the foundation) know which website (also called *relying party* or *verifier*) you share attributes with. This is also referred to as *issuer unlinkability*. An IRMA server is required to perform IRMA sessions with IRMA apps. It handles all IRMA-specific cryptographic details of issuing or verifying IRMA attributes with an IRMA app on behalf of a requestor (the application wishing to verify or issue attributes). Please note that the IRMA server might have access to the attributes. To avoid this protentional big brother, the foundation proposes that each verifier and issuer runs its own instance of the IRMA server, contrary to this centralized IRMA server.

---

[6] An earlier version of this chapter was review by Sietse Ringers from the Privacy by Design foundation, and a later version by Bart Jacobs (Privacy by Design foundation and Radboud University). Of course, errors or mistakes and opinions remain the responsibility of the authors, not of the reviewer.

[7] `https://www.zurich.ibm.com/identity_mixer/`

[8] From the "about IRMA" section from the Privacy by Design foundation.

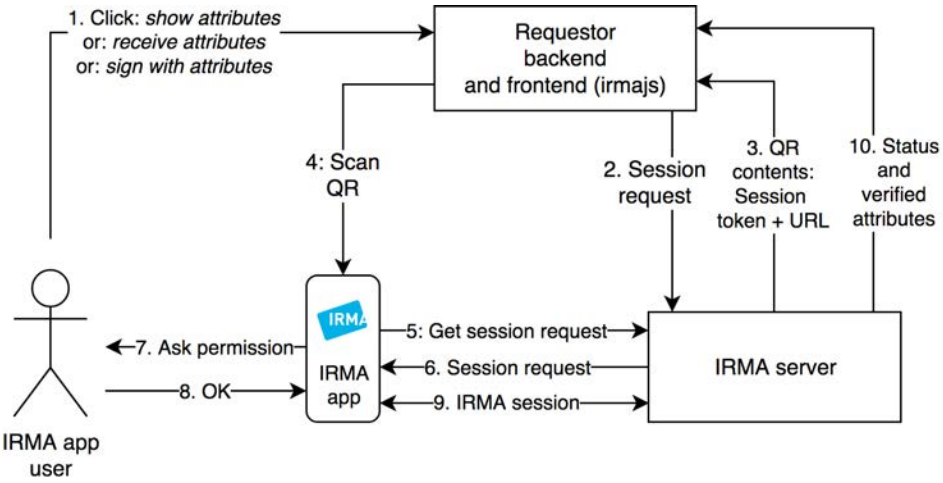


Figure 2 Typical IRMA flow (source: IRMA). The requestor can either be the issuer or the verifier.

2.1.3 Provided issuers by the Privacy by Design Foundation

The trust in the attributes available in IRMA is directly linked to the issuers of the attributes, i.e., it is never better than the issuer of the attributes. The architecture of IRMA allows adding new issuers, but currently the for this report relevant issuers of IRMA attributes are:

- **Gemeente Nijmegen/BRP/DigiD** – Via Gemeente Nijmegen to the Dutch central administration for civilians (i.e. basic registration of persons, BRP), after a login via DigiD. These attributes include address, age limits, BSN and personal data (name, gender, if Dutch or not). Although the attributes are from the BRP, they are issued by Gemeente Nijmegen. They are categorised in a few credentials.
- **iDIN** – which is also part of the remote vetting PoC. iDIN attributes are address, name (initials and last name), gender and date of birth. The attributes originate from the user's bank. The issuer is the Privacy by Design foundation and consists of two credentials: (i) name, age and address and (ii) derived age limits.

The details and complete list of issuers can be found here: `https://privacybydesign.foundation/attribute-index/en/` and `https://github.com/privacybydesign/pbdf-schememanager`. This also includes SURF, which participates as an issuer of the SURFsecureID credential, i.e., someone who has done the step-up on their SURFconext account with SURFsecureID can store this as a credential in the IRMA app.

With respect to BRP/DigiD, this is done as a pilot[9] via the Gemeente Nijmegen[10], originally for a max 7.500 persons per month but this was extended later on. Gemeente Nijmegen requires the user to use DigiD SMS or mobile app for logging in at the BRP (i.e., *DigiD Midden* trust level). The issuer from the foundation (or scheme) perspective is the Gemeente Nijmegen, and not e.g. Logius or RvIG, even though the actual citizen may live in another city than Nijmegen. This seems to be a precedent for the Dutch government, that a private organisation (the Privacy by Design foundation) is, indirectly via a municipality and the citizen in question, effectively is allowed to process BRP-attributes including user's citizen service number (BSN). The motivation of the Gemeente Nijmegen to do this is twofold: be GDPR compliant and give the user more control over its personal attributes. In this case, the Gemeente Nijmegen runs its own IRMA server, the foundation does not get access to the personal data. For the BRP attributes, the Gemeente Nijmegen is the issuer and provides the attributes using the IRMA/Idemix crypto. In general, it is preferable for the authoritative source to be the issuer, and not the foundation. That a government organization is the issuer, in combination with that it is

[9] `https://www.nijmegen.nl/nieuws/app-irma/` (18 Dec 2018)
[10] `https://privacybydesign.foundation/uitgifte-brp/` (18 Dec 2018)

currently not possible for SURFnet to directly use BRP/DigiD, makes IRMA/BRP an attractive option to include in the SURFsecureID PoC. For iDIN however, the foundation access, signs and using iDIN via IRMA therefore significantly reduces the trust and adds complexity to the customer journey without adding much value.

The foundation and the city of Amsterdam indicated that more city counsels' will be joining the pilot[11]. This precedence may help others to also get this access, and possibly SURFnet can also use BRP/DigiD for remote onboarding without involvement of the foundation, and thus the scheme and app of the foundation, however they would have to collaborate with a municipality to do so. If the white label IRMA platform is successful, then this needs to be verified.

An important characteristic of IRMA is that attributes can be shared on a per-attribute basis. In IRMA terminology, the user reveals only those attributes he wants to reveal. The per-attribute sharing, the decentral architecture and the above-mentioned issuer unlinkability are important privacy features of IRMA. Important to realize is that there are two ways in which the foundation can pass attributes from the trusted sources to the Verifier. It can simply download the attributes and sign them, i.e., the foundation becomes the issuer, therefore breaking the trust chain (i.e., the verifier cannot verify if the attributes are actually for the claimed trusted source, as is currently the situation for iDIN attributes) or if the issuer uses IRMA/Idemix specific crypto it can pass the attributes without re-signing / breaking the trust chain (as done for BRP).

A second important characteristic of IRMA is that the received attributes are stored, or cached, on the phone. They may change, or be revoked, by the original source, but this will currently not impact the stored credentials on the phone. In addition, a compromised or 'borrowed' DigiD or bank-account may be used to load the attributes to the app, but then these attributes cannot be revoked from IRMA at the moment[12]. This is a consequence of the decentral and privacy-by-design architecture of IRMA, in which it is basically not known which attributes are present on what phone, so these cannot be revoked[13]. Attributes will expire after a certain period, i.e., incorrect attributes can only be used for a certain period, after which a user has to re-obtain the attributes. Attributes are also time-stamped with the date they were issued by the source[14].

A third characteristic of IRMA is the binding between the IRMA and the user. The user can revoke a stolen phone, and with that all the loaded attributes. For this the user needs to have linked his phone to his MijnIRMA environment. The IRMA app is protected by a pin code which is common practice for mobile based personal data sharing solutions.

The business model of the Privacy by Design foundation, and thus the costs for SURFnet, is still somewhat open. The current idea is that it is free for the user and verifier (which is good for the SURFsecureID remote vetting use case), but this may not be a durable business model for the foundation. For the PoC the long term business model is not a pressing concern, but if the PoC is successful then this may be worth exploring.

2.2 IRMA analysis

2.2.1 Assessment criteria

Remote vetting solutions have to fulfil to a number of criteria. These criteria are derived from interviews and discussion sessions with SURFnet and stakeholder institutions, executed in the previous research project.

[11] IRMA meeting 5 july 2019.
[12] IRMA plans to add this.
[13] Revocation by the user himself is possible, e.g., if a phone is lost, but the treat scenario here is a malicious user loading attributes from someone else on his phone. IRMA plans to add this in the future.
[14] See `https://credentials.github.io/docs/irma.html#special-attributes`.

Criteria for remote vetting are:

- **Easy to use by the user**: if the user experiences inconveniences during remote vetting he may cancel the process. For instance, many users would like to be able to obtain a SURFsecureID token outside office hours. Compared to the current practice, the ease of use of the solutions from a user perspective can be either better, equal or worse.
- **Easy to organize by the institution**: it must be easy for the institution to enrol, deploy, initiate, or arrange a remote vetting solution. Compared to the current practice, the ease of use of the solutions from an institution perspective can be either better, equal or worse.
- **Limited impact on current SURFsecureID service**: how easily can the remote vetting solution be integrated with the current SURFsecureID service, what needs to be adapted technically or organisationally by SURFnet, is it a one-off (e.g. software improvement) or continuous (e.g. audit process) effort? Solutions have no, limited or large impact on the current SURFsecureID service provisioning.
- **Straight-through processing**: the possibility to vet for the user's identity in a fully automated manner without human interference. More automation means shorter vetting lead times and improves the user experience. It also provides more efficiency, scalability and less errors (e.g. due to manual processing of personal information). For bulk enrolment scenario's this is very relevant. The automation capabilities of the vetting process are less, similar or better than the current situation offers.
- **Sufficient penetration rate**: as many potential target users as possible must be able to go through a remote vetting process. Certain user groups may not be able to execute the remote vetting process because they lack certain functionality that is required for remote vetting (e.g. they use a phone that does not support NFC or do not have a Dutch bank card). The penetration rate is higher, equal or lower compared to what the existing SURFsecureID solution for vetting.
- **Sufficient level of authentication assurance**: the outcome of the remote vetting must provide sufficient assurance in the identity of the user (which on its turn will provide a higher authentication assurance). Solutions must achieve a level of assurance that at least correspond to LoA 2 and LoA 3 as defined by SURFsecureID.
- **Costs**: the cost of the solution is reasonable. The current service desk costs are estimated to be about a minimum of 5 Euro per vetting15. User costs are also involved. However, these are more difficult to quantify as the costs for students are different than for employees. Therefore, the user's costs are taken into in the ease of use criterion above. There are other costs as well, such as development costs (only once), licensing costs (recurring), authentication costs of iDIN if used as an issuer, and technology/hardware costs. However, these costs are expected to be similar for all solutions. The focus therefore will be on the costs for vetting the user. Consequently, the costs assessment of remote vetting solutions will be rated as higher, similar or lower than 5 Euro. Specific, significant additional costs will be mentioned during the assessment if needed.
- **Controllability/auditability**: the ability to control the remote vetting process in such a way that it is implemented by all institutions in an unambiguous manner including the ability to audit the process for accountability purposes. The controllability/auditability of remote vetting solutions is better, similar or worse than what SURFsecureID currently offers.
- **Future proof**: Is the solution future proof and does it have a sufficient maturity level?

2.2.2 Assessment against criteria

We fill this in based on the current sources for attributes, i.e., iDIN and DigiD/BRP.

Table 1: IRMA/PbD foundation assessment

[15] The costs are estimated as follows: on average it takes a service desk employee about 6 minutes to verify the user's identity and to activate the token. This employee costs the institution about 50 Euros per hour. So the costs of a single vetting amount to 5 Euro.

Criteria	Assessment	Score
Easy to use by user	Easy to use. The attributes are loaded on the IRMA app, relevant are DigiD/BRP (and to a lesser extent, iDIN). Score very much depends if the app is also installed with valid credential or not.	3
Easy to organize by institution	Similar as iDIN, ReadID/NFC app etc	5
Limited impact on SURFsecureID	Same as other apps	3
Straight-through processing	Yes	5
Penetration rate / coverage	This is de-facto a pilot, currently no significant coverage. But users can install the app and load the attributes, the attributes have a very good coverage (DigiD/BRP and iDIN) of all inhabitants in the Netherlands. Only Dutch sources though. Works on iOS and Android devices.	3
Assurance level	The assurance level is less than that of the original sources, since the data may be outdated and the original signature from the issuer is broken (in case of iDIN), i.e., no end-to-end trust chain. DigiD SMS or mobile app is required to get access to BRP.	3
Costs	Free, but long-term business model is unclear at the time of writing.	5
Controllability/auditability	It is open source, but the foundation itself has not been audited (yet).	3
Future proof	No clarity on longer term durability.	3

Total score is 33, slightly lower than the iDIN and ReadID scores (39).

2.3 scoring use cases

There are four typical use cases for remote vetting, each of which poses its own unique set of requirements to the remote vetting process.

The first use case involves a small target group of users at an institution that does not have a Registration Authority for physical registration. In that sense these users aren't necessarily remote users.

The second use case involves a relatively small target group of remote users that cannot visit the RA of the institution. Users will typically be Dutch researchers or employees that live and work in or outside The Netherlands.

The third use case involves a relatively small target group of remote users that cannot visit the RA of the institution. Users will typically be foreigners that live outside The Netherlands.

The fourth use case involves the identity vetting of large amounts of users via a remote solution; i.e. bulk enrolment.

In the table below the assessment of IRMA against the use cases is described.

Table 2: IRMA use case assessment

Use case	Assessment	Verdict
1 Small amount of users	With IRMA users can self-service their vetting process and do not need an RA at their institution.	
2 Remote Dutch users	All Dutch citizens are registered in the BRP and most likely have DigiD, regardless of whether they live in the Netherlands or abroad. This makes IRMA a suitable solution.	
3 Remote foreign users	Using IRMA to obtain attributes from the BRP requires foreigners to be registered in the BRP and therefore have a BSN. This may be the Non-residents Records Database. To access the BRP, they also need a DigiD. If foreign users are formally employees of Dutch institutions, they are registered in the BRP and likely have DigiD. If, however foreign users are not somehow connected to the Netherlands they will not be registered nor have DigiD. This makes IRMA unsuitable for foreign users.	
4 Bulk enrolment	Technically, IRMA is suitable for bulk enrolment. It is currently unclear how this situation will develop in the future.	

1.1 Conclusions and reccommendations

IRMA relays cached attributes from trusted issuers in a privacy-friendly manner. A major benefit of IRMA is of course that unneeded, and thus unwanted, attributes are not shared with SURFsecureID. And the transparency of the whole process for the user will likely appeal to privacy-concerned users.

The trust in the attributes can never be higher than the trust in the actual source, e.g., the Dutch BPR, and the authentication method that was used to 'load' the attributes in the app. For the remote vetting use case, the by far most interesting source of attributes provided by the Gemeente Nijmegen, which gives access to the BRP data of all Dutch citizens, also from other municipalities. The attributes are loaded in the IRMA app after the user logs in using DigiD Midden. BRP is considered a trusted source of attributes, and the Gemeente Nijmegen is a trusted issuer even though having a specific municipality as issuer is not the obvious issuer for the whole BRP (this may change). The authentication that the Gemeente Nijmegen relies on however is DigiD Midden,

i.e., username/password + SMS (or DigiD App), which means that although the attributes themselves are authentic they may belong to a different person since DigiD Midden has an eIDAS Low trust level. This is a risk to consider in the overall assessment of the trust. Increasing the authentication level to Substantial would increase trust significantly, but has coverage disadvantages (no iOS availability, similar to ReadID), and to a lesser extend also usability disadvantages.

A consequence of caching is that the attributes may get outdated. This in itself is more a theoretical risk for the specific remote vetting use case, since the needed attributes are unlikely to change. There is, however, a second risk associated with the caching. If, at the moment in time that the attributes were 'loaded' in the IRMA app, the credentials used by the user were compromised, the attacker can use those attributes until they become outdated. Due to the privacy features of IRMA, even when known there is an attacker out there with someone's attributes, it is currently not possible to revoke those attributes. The user can however revoke his complete phone, with it all its attributes. Requiring the attributes to be valid only for a very short period mitigates that risk somewhat but will effectively require the user to load the attributes when doing the remote vetting. This reduces the user friendliness compared to simply sharing pre-loaded attributes of, e.g., two months ago. Since the attributes are time-stamped, SURFnet can define a policy that limits the validity period of the IRMA attributes required for remote vetting.

A related weakness of IRMA is that the loading of attributes via a PC is performed by scanning a QR code (when done via a mobile phone this weakness does not exist). The attribute source displays this code on a webpage, the user scans it with the IRMA app. For the party displaying this QR code, there is no way to verify the QR code is being scanned by the right person at the moment of scanning. Someone that has access to the QR code that is displayed can 'steal' those attributes, and present those as his own. These attributes will have to expire. Furthermore, this means the binding between the user and the IRMA app is weak. The sharing of the attributes from IRMA to SURFsecureID is done in the same way, by scanning a QR code.

Although IRMA positions itself as an alternative to a broker, we could see IRMA as quite a similar intermediate, but then in a more privacy-friendly manner. There are, however, disadvantages to this more privacy-friendly and decentral architecture, especially that this can break the trust chain. This however depends on the issuer of the attributes. If the Idemix/IRMA protocol is implemented, then an end-2-end trust chain remains in existence. Currently, for BRP/Gemeente Nijmegen the end-2-end trust stays. However, for both iDIN attributes and derived attributes (like 65+) it does not; i.e., iDIN attributes that SURFsecureID would receive are signed by IRMA/foundation and not by the bank. In other words, one has to trust the Privacy-by-Design Foundation for these identity assertions. Moreover, the Privacy by Design foundation is currently not subject to external audits/checks, i.e., although it is run by people with an excellent trust reputation, trusting attributes signed by an intermediate third party that is not audited on its security practice has an inherent risk. There are thus no compelling reasons to use iDIN attributes via IRMA/foundation. It may be more convenient, in the sense that SURFsecureID only needs to connect to IRMA/foundation (i.e., run an IRMA server). This however entails a dependency on IRMA/foundation and leaves the users with no choice but IRMA/foundation. Furthermore, it unnecessarily lengthens the chain of communications lowering the reliability. It is thus favourable to use iDIN directly, as is also part of the PoC. Only for cost reasons one may opt for IRMA as it is currently for free even when iDIN is used as an issuer.

Should access to BRP attributes via the Gemeente Nijmegen continues to be possible (or via another government organization) beyond the current pilot, then by extension possibly SURFsecureID can access without involvement of the foundation. This would make an interesting option.

Overall, the access to BRP using DigiD that Gemeente Nijmegen provides via the foundation is interesting and including it in the PoC and subsequent pilot allows SURFnet to explore the value this may bring.

3 Matching Analysis

In the remote vetting process external identities from iDIN, ReadID or IRMA are to be matched with the identity provided by the institution's IDP in order to increase the assurance level of the user's identity. We have analysed the identity assertions of all providers (iDIN, ReadID, IRMA/BRP and institutional IDP) to be able to determine the accuracy and reliability of the matching process. The approach and outcomes are described in this section.

3.1 Goal

Goal of the matching analysis is to analyse the quality of the matching of identity attributes from the institutions with attributes from iDIN, passport chips (ReadID) and IRMA. This includes analysing the consequences for the reliability of the identities compared to the current process.

Specific research questions to be answered are:

1. How do the institutions deal with maiden names?
2. How do the banks / iDIN deal with maiden names and names in general; i.e. full first names or initials only?
3. Are there any additional attributes present at banks / iDIN that can help improve the reliability of the matching, such as gender and date of birth?
4. How do institutions deal with date of birth? Is date of birth available for their identity provider? Are they willing to release it for remote vetting?
5. How thorough is the verification of the attributes, for both the institutions as well as the different remote vetting means?
6. What are the differences in attributes for the different types of users (students, employees, flex workers, etc)?
7. What false positive and false negative rate is expected for the matching?

3.2 approach

A limited group of users (8, friends and family) was asked to perform an authentication session with two or more solutions. Via SURFconext's debug page[16] users were able to authenticate with iDIN and/or the institutional IDP. For ReadID and IRMA they had to install the respective apps on the mobile phone. The outcomes of the authentication sessions were aggregated in an Excel sheet. For privacy reasons, not all personal information was processed, and other information was anonymised.

3.3 results

The table below summarises all the identity attribute assertions for the various authentication solutions. For ReadID we separate passports[17] and driving licenses.

[16] https://engine.surfconext.nl/authentication/sp/debug
[17] Identity cards will have equal attributes to passports.

Table 3: Identity attributes as obtained with iDIN, IRMA/BRP, ReadID passport and driving licence, and SURFconext IDP

Attribute	iDIN	IRMA	ReadID passport	ReadID driving license	SURFconext IDP
Full name	1. Doornbosch, RJ (interpreted) 2. Doornbosch - Olfen, HA (interpreted) 3. - 4. - 5. - 6. – 7. – 8. van den Burg, ECR (interpreted)	1. R.J Doornbosch 2. - 3. P.G.M. van der Molen 4. L. Klaas 5. M. Wegman 6. – 7. -	1. Doornbosch, Robertus Johannes 2. Olfen, Helena Anna 3. - 4. Klaas, Lara 5. – 6. – 7. -	1. Doornbosch Robertus J 2. - 3. - 4. Klaas Lara 5. – 6. – 7. -	1.- 2. - 3. Peter van der Molen 4. - 5. - 6. T.F. Duits (UT) 7. Toon Hoeks (TUD&Saxion)
First name(s)	1. - 2. - 3. - 4. - 5. - 6 - 7. – 8. -	1. Robertus Johannes 2. - 3. Peter Gerrit Martijn 4. Lara 5. Mats 6. – 7. -	1. Robertus Johannes 2. Helena Anna 3. - 4. Lara 5. – 6. – 7. -	1. Robertus J 2. - 3. - 4. Lara 5. – 6. – 7. -	1. - 2. - 3. Peter 4. - 5. - 6. Toosje 7. Toon (TUD&Saxion)
Surname	1. Doornbosch 2. Olfen 3. Molen 4. Klaas 5. Wegman 6. – 7. – 8. Burg*	1.- Doornbosch 2. - 3. Molen 4. Klaas 5. Wegman 6. – 7. -	1. Doornbosch 2. Olfen 3. - 4. Klaas 5. – 6. – 7. -	1. Doornbosch 2. - 3. - 4. Klaas 5. – 6. – 7. -	1. - 2. - 3. van der Molen 4. - 5. - 6. Duits 7. Hoeks (TUD&Saxion)
Initials	1. RJ 2. HA 3. PGM 4. L 5. M	1. R.J. 2. - 3. P.G.M. 4. L. 5. M.			

	6. – 7. – 8. ECR	6. – 7. -			
Date of birth	1. YYYYMMDD 2. YYYYMMDD 3. YYYYMMDD 4. YYYYMMDD 5. YYYYMMDD 6. – 7. – 8. YYYYMMDD	1. DD-MM-YYYY 2. - 3. DD-MM-YYYY 4. DD-MM-YYYY 5. DD-MM-YYYY 6. – 7. -	1. YYMMDD 2. YYMMDD 3. - 4. YYMMDD 5. – 6. – 7. -	1. YYMMDD 2. - 3. - 4. YYMMDD 5. – 6. – 7. –	
Gender	1. 1 2. 2 3. 1 4. 2 5. 1 6. – 7. – 8. 2	1. M 2. - 3. M 4. V 5. M 6. – 7. -	1. Male 2. Female 3. - 4. Female 5. – 6. – 7. -		
Nationality	1. NL 2. NL 3. NL 4. NL 5. NL 6. – 7. -	1. Ja (Dutch) 2. - 3. Ja (Dutch) 4. Ja (Dutch) 5. Ja (Dutch) 6. – 7. -	1. NLD 2. NLD 3. - 4. NLD 5. – 6. – 7. -		

* The last name of the partner is sometimes also provided by iDIN.

3.4 Analysis

A few things from this small experiment stand out.

3.4.1 First name(s) and initials

iDIN does not provide first names; only initials. IRMA and ReadID passport provide the full first names. ReadID driving license only provides the full first name and switches to initials for additional first names. The IDP only provides the full first name. Only iDIN and IRMA provide initials.

IRMA and ReadID passport/driving license provide the first full name and this matches well with the approach several IDPs seem to have taken. IDPs that have adopted an initials approach can also be relatively easily matched with these solutions.

Since iDIN does not provide full first names, when matching with IDPs that provide only the first full name it will be challenging to obtain adequately low false acceptance and rejection rates. Particularly if additional information such as date of birth is not available. For the IDPs that provide initials, the matching is easier, but there will be uncertainty if the same user is matched.

To get an impression of the extent to which a certain set of attributes delivers a unique hit in the Dutch population register (BRP), a query was run on a representative set of data[18]. When searching for full first names, surname and date of birth, more than 5,000 couples were found that meet the identification. In other words, there are several persons with the same identifying attributes. When searching with initials, surname and date of birth (the first names will not usually be offered in full), around 30,000 couples were found. This number is even higher if use is made of 'intelligent search' algorithms that abstract from e.g. diacritical marks, prefixes and similar looking names (e.g. Janssen vs Jansen). See also Section 3.5.1.

3.4.2 Surname

The issue with surnames is prefixes: the IDPs seem to differ here from the other providers by including prefixes in the surname attribute.

3.4.3 Full name

How the various providers make use of the full name attribute is quite messy. iDIN does not provide full names, only interpreted ones as a combination of last name and initials and sometimes with the spouse's surname. IRMA and ReadID passport provide the full combination of first names and surname. ReadID driving license only provides the first full name, initials for the other first names and the surname. IDPs seem to vary in their full name strategy. The table above shows two examples: first full name and surname or initials and surname. But other full name combinations have been reported as well:

surname, prefix and first name: Meulen, van der Pieter;

first name, abbreviated prefix and surname: Pieter vd Meulen;

initial, prefix and surname: P. van der Meulen.

3.4.4 Date of birth

Regarding date of birth various formats are used:

iDIN: YYYYMMDD

IRMA: DD-MM-YYYY

ReadID passport & driving license: DD.MM.YYYY

IDP: not provided.

Matching between the various solutions can be easily achieved and implemented via a translation function. Unfortunately, most IDPs do not provide this attribute. It is recommended that they will do so for

[18] Source: Use cases eIDAS – BRP, Frans Rijkers, Rijksdienst voor Identiteitsgegevens, 2 april 2015, Werkdocument t.b.v. de werkgroep eIDAS.

remote vetting purposes because it allows the remote vetting operator to (more) uniquely identify the user and to be able to match and link the user's electronic identities with more assurance[19].

3.4.5 Gender

Unexpectedly, the use of the gender attribute differs quite a lot across the various providers. iDIN provides 1 or 2 for male or female, IRMA provides M or V, ReadID passport provides Male or Female[20]. ReadID driving license and the IDPs do not provide the gender attribute. Despite the variation, matching can be done quite easily with a simple translation table.

3.4.6 Address

Address information is only provided by iDIN. This may be useful in case a second factor authentication token has to be sent to the user via regular mail. For privacy reasons, address is left out of the above table.

3.4.7 Nationality

The nationality attribute is only provided by iDIN, IRMA and ReadID passport. The value of the attributes is either NL (for iDIN) or Dutch (for IRMA and ReadID).

3.4.8 Reliability of the attributes

Another aspect to take into account is the reliability or accuracy of the attributes provided by the various identity providers. Here ReadID probably scores best as it provides identity attributes that are read from the chip of a valid identity document during the remote vetting process. For IRMA, the identity attributes provided have been previously obtained from the BRP and can be up to 90 days old. The timestamp of the attributes can be used to decide whether or not to accept the identity attributes from IRMA and to force the user to upload fresher ones. This is solely relevant for the Surname attribute as the other attributes are unlikely to change. Also note that the other two solutions (i.e. iDIN and ReadID) face the same problem. The reliability of the identity information provided by iDIN is shown in Figure 3 (in Dutch) proves that attributes in general are pretty static.

Meetcriteria	Uitleg	Norm
Accuraatheid	De mate waarin iDIN-datavelden overeenkomen met hetgeen er is vastgelegd in de kopie identiteitsbewijs.	95 – 99 %
Compleetheid	De mate waarin iDIN-data aanwezig is: de datavelden uit het datamodel zijn gevuld.	98 – 100 %
Correctheid	De mate waarin iDIN-data voldoet aan het vereiste veldformaat: voldoet aan de formatting rules zoals beschreven in de meest recente versie van de implementatiegids iDIN. Een attribuut moet kunnen worden geleverd conform de formatting rules.	99 %
Uniciteit	De mate waarin iDIN-data (i.c. BSN) uniek is: slechts één keer voorkomt bij een Issuer.	97 %

Figure 3: iDIN accuracy, completeness, correctness and uniqueness of identity information [source: iDIN product sheet 2017, see `https://www.idin.nl/cms/files/Productsheet-iDIN.pdf`].

[19] In addition: it is one of the required attributes of the European eIDAS regulation for electronic identification. Being compliant with eIDAS, will lead to better eID interoperability across Europe.

[20] Legal identity documents can have "X" as gender specification, though this will only occur very sporadically.

3.5 impact on remote vetting process

What do the outcomes of this little experiment mean for the remote vetting process?

3.5.1 Matching challenges

Generally speaking, matching identities from different systems and with different formats is not easy. Here are some common matching challenges to be dealt with when matching identities:

- Diacritical marks (à á â ã ä ā ă ė ä å ç ő ą ě) are removed in certain systems (e.g. the Machine Readable Zone of identity documents does not contain diacritical marks);
- Special characters (Æ æ Đ đ Ħ ħ ı ĸ Ŀ· l· Ł ł Ŋ ŋ ’n Ø ø Œ œ ß Þ þ Ŧ ŧ IJ ij) are translated similar to the ICAO-rules for the Machine Readable Zone;
- Uppercase characters are replaced by lowercase characters;
- Every other character than a..z or 0..9 is replaced by <space>;
- All <spaces> are removed;
- Phonetic equivalents are replaced (longest first):
    - v w
    - a ae
    - tsch sch tch tsj zj zh sh ch sj jh kh x s
    - schtsch sjtsj schch chtch sc
    - ij and y
    - oe ou yu ue o u
- Multiple same adjacent characters are replaced by one character;
- Remaining characters h are removed.

Note that these translations may differ per context (i.e. iDIN, IRMA, ReadID, IDP). The origin of these matching issues is diverse: the variety of systems that process the personal information differently, standardisation requirements concerning the format of a e.g. the MRZ, and not being aware of the consequences a small change of personal data has for its further processing. For example, the source identity information of persons applying for a visa is the MRZ of a legal identity document (i.e. passport). Since the MRZ does not contain diacritical marks, the person’s identity data on the visa will be different compared to the data on the passport and its chip.

The challenges with diacritical characters depend very much on the language in which the name is formed. English names hardly cause any problems. French a bit more, but because those diacritics have always been in ASCII, that often goes well too. With Polish names it is the Polish ł that one should pay attention to. With transliteration and transcription – the names that originally stood in a different script such as Greek, Cyrillic, Chinese – more things can go wrong. When converting to Roman script, it depends on who does it, and also in which country that happened. If the conversion has always been done in the original country, there is a good chance that it has been converted in the same and correct way each time. But if, for example, a Romanian has lived in Germany or France for a while and then comes to the Netherlands, there is a good chance that the name will be different compared to the original one.

The biggest challenge is reducing the chance of wrong matches, i.e. matching the identity of a user to the that of somebody else. This includes when an attacker attempts to exploit this weakness to induce a wrong match. It is therefore important to find a balance between providing a service with just the elements ‘family name’ and ‘date of birth’ versus no wrong connections. Finding this balance is not trivial. This is illustrated by the following example of the Dutch personal data register[21]:

[21] Source RvIG.

- Number of identities in the register: 21 million;
    - 17 million residents + 4 million non-residents (+ 3 million deceased);
- Spread over 22.000 birthdates (= 60 year);
- Means 1000 identities per birthdate on average;
- Chance of finding more than one identity with the combination Family name + Birthdate:
    - 40 family names have a frequency of 1+ per thousand;
    - 15 of them have a frequency of 2+ per thousand;
    - 7 of them have a frequency of 3+ per thousand;
    - More concrete:
        - Jan(s)sen 8 per 1000 Jan(s)sen
        - De Jong / De Vries 5 per 1000 De Jong / De Vries-en
        - Vd Berg / van Dijk / Bakker 4 per 1000 Bakkers
        - Visser 3 per 1000 Vissers
    - Taking into account 'gender' will alter the statistics by ~50%.

So, for certain identities, there is serious risk of a false match. The NIST Special Publication on identity assurance recommends that the matching should at least be better than 1 in 1000 for biometric authentication[22]. This means that when 1000 users try to authenticate biometrically, one of them is accepted under another identity. A similar situation may arise when doing identity matching for e.g. J. Jansen. Unfortunately, NIST does not specify for which assurance level this rate is applicable, i.e. is 1/1000 acceptable for High or Substantial?

Adding more attributes to the matching algorithm helps, but may be at odds with privacy legislation. This is another challenge. It may be worthwhile to consider executing a privacy impact assessment on the matching service. Special attention in this case is needed for the balance between service being provided and risk management: false positives vs false negatives vs fraud prevention/detection. E.g. a mismatch may lead to the wrong user accessing personal details of another user and consequently this may lead to a data breach that needs to be reported to the Data Protection Authority. To prevent privacy issues, it is recommended to store the matching data for a limited period, i.e. for the duration of the vetting process and delete the data after e.g. one month.

[22] See `https://pages.nist.gov/800-63-3/sp800-63b.html`.

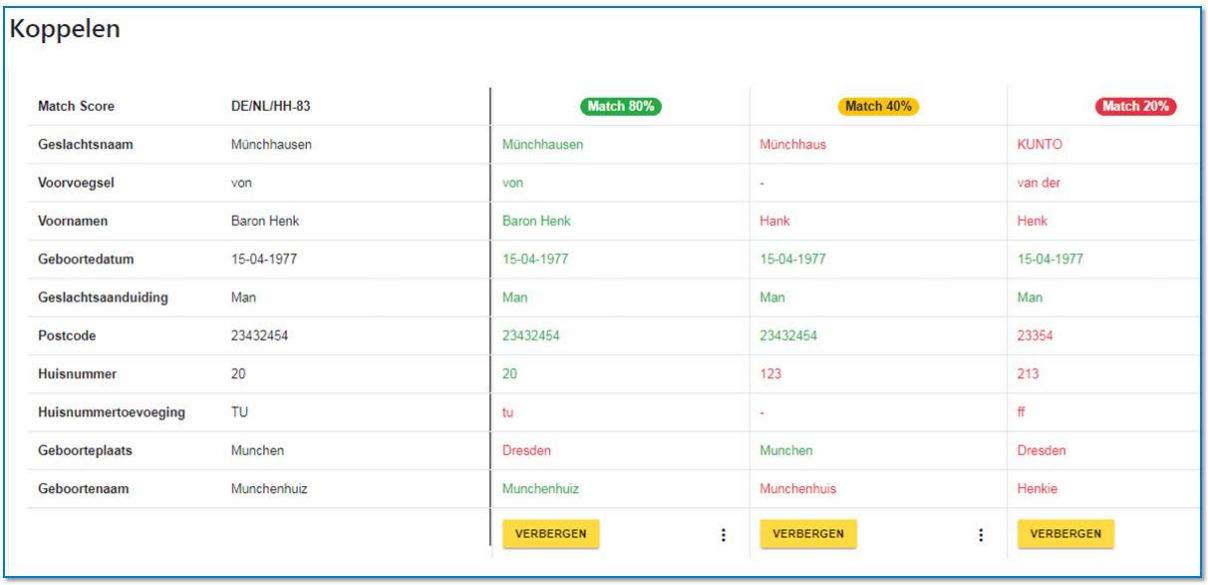


Figure 4: Matching example eIDAS BRP matching service [source: Identity matching presentation by Frans Rijkers for a seminar on eIDAS and CEF in Brussels on 29th January 2019].

Experiences gained from Idensys and eHerkenning for the BSN coupling registers show that matching can be quite difficult due to the above-described reasons. For some providers it was not possible to match an identity with the Dutch Person Register in 10% of the cases. BKR experiences similar challenges for applications made by citizens. To summarise:

- Identity matching is not trivial and it is difficult to express the probability of a correct match in a single statistic as this may vary per family name.
- Translations on the side of the identity providers may differ and should be normalised prior to matching.
- Matching without date of birth is almost impossible and will likely result in too many false acceptances (this is confirmed by experts of BRP).

3.5.2 Matching strategy

In order to mitigate the risks of remote vetting and matching the following matching strategy could be adopted, in which each stage will only be executed if the stage before fails. See Figure 5. Starting point of the strategy is to keep it simple: start with basic matching of strings of personal data without fancy or fuzzy logic rules. We propose the following steps:

- **Stage 1**: Attribute matching based on attributes provided by the institutional IDP and the external source (i.e. iDIN, ReadID or IRMA). Matching based on full name if available or on initials (in case of iDIN), date of birth and, if available, gender. E.g. if Lara Klaas and L. Klaas are provided by the institutional IDP and iDIN respectively, the matching algorithm will be to only use the initial of the user's first name asserted by the IDP and compare that with the iDIN assertion: L. Klaas vs L. Klaas. In this case there will be a match.
- **Stage 2**: The user gets the opportunity to try a second remote vetting method, e.g., if matching based on iDIN failed, then the user can try IRMA/BRP. Matching is performed again based on full name if available or on initials (in case of iDIN), date of birth and, if available, gender. The user can choose to not try a second remote vetting method, e.g., when matching based on ReadID attributes failed and the user does not have a Dutch bank account and DigiD. Initially, the identity attributes provided by the second external identity provider will be matched with those provided by the institutional IDP. If there is a match, the remote vetting process will continue. In case there is no match, a second matching attempt will made. This time, the identity attributes provided by both external identity providers (e.g. iDIN and IRMA) will be matched.

In case of a match the remote vetting process will continue; in the absence of a match the RA will be involved (see next stage).

- **Stage 3**: The RA assesses if the attributes match, expanding on the automated matching of the previous stages. RA can e.g. also involve HR department, do some other form of gathering additional evidence to determine if there is a match. If there is no match, then the RA indicates as such in the portal, and can send the user a message, e.g., contact your HR department to correct a possible wrong date of birth or to go to the physical registration process.

Obviously, it is recommended to regularly audit matching requests and their outcome to help refine the matching strategy. The above three stages mirror the existing process (with a more-or-less fuzzy human matching step by the RA if the automated matching fails), and the retry in stage 2 we expect will reduce the amount of users that end up with in stage 3.

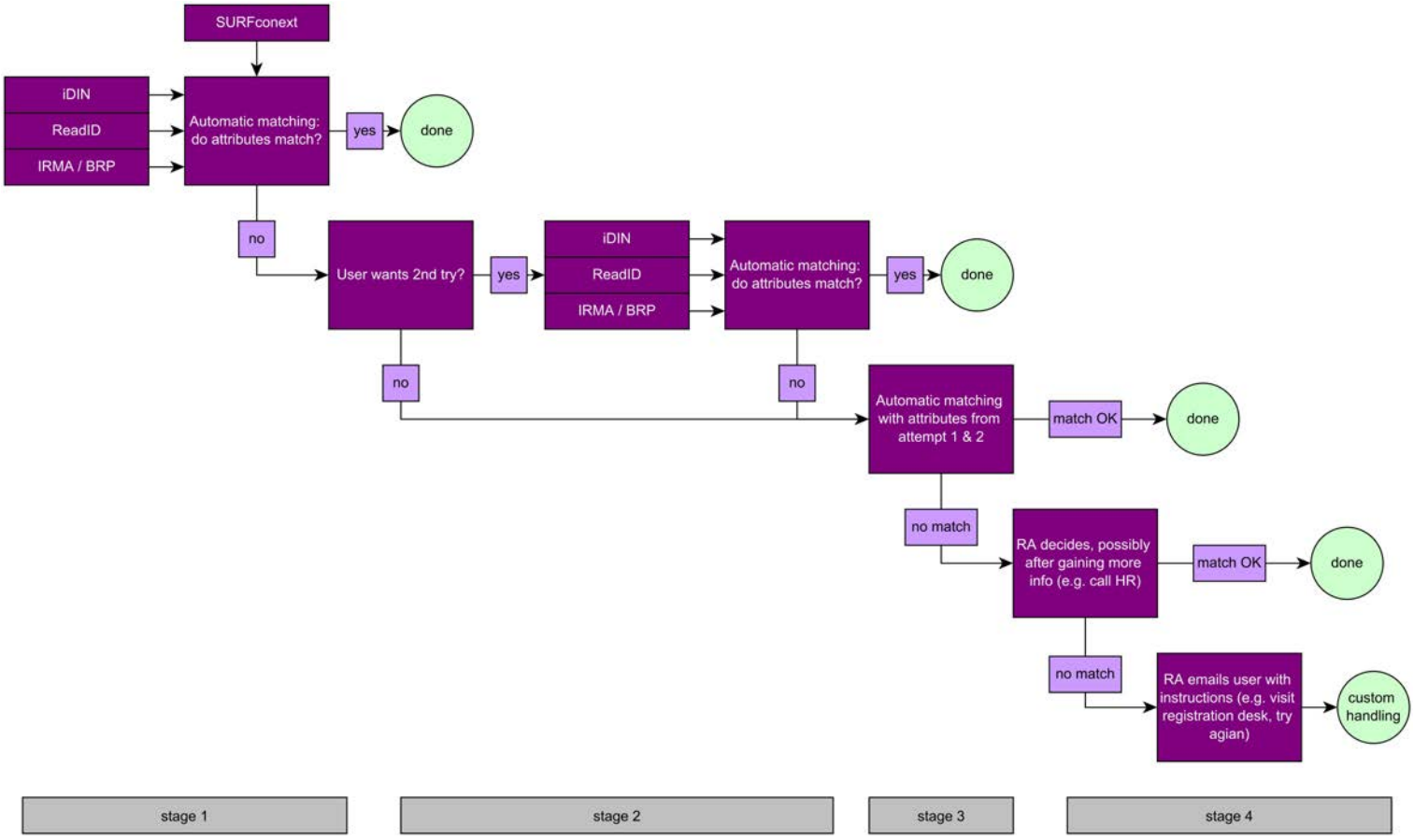


Figure 5: Matching strategy in three stages

3.5.3 Impact of (mis)matching on the remote vetting process

What will be the impact of these matching challenges on the remote vetting process? Obviously, there is a real chance of a mismatch. This means that someone's institutional account could very well be coupled to the identity information provided by iDIN/IRMA/ReadID of another person (accidentally or on purpose by an identity fraudster). The impact of such a mismatch in the remote vetting process is a risk that needs to be mitigated.

Consider the following example of how the identity matching process could be exploited by an attacker. A malicious user could relatively easily hack (e.g. using password phishing, guessing or stealing) the institutional account of a victim because this is based on single factor authentication. This is exactly the risk that adding a second factor authentication addresses. Having hacked the institutional account, the hacker either needs someone with a bank account on a name similar to that of the victim (e.g. by using a 'mule'), or have a (stolen) identity document of the victim. Due to poor matching statistics, the matching will succeed. This allows the hacker to obtain a 2nd factor token that is required to access the critical/sensitive services that are protected by 2FA. Note that these attacks scale very poorly, so there must be a very strong incentive for an attacker to spend all the effort to compromise the user's account in order to add an additional authentication token.

Moreover, in the current process of user identification at the service desk of the institution, the RA has to deal with similar matching issues. The RA has to match the attributes provided by the IDP with those on the identity document shown by the user at the desk. On the other hand, the physical context in this case may discourage an attacker to try to get a second authentication factor on someone else's identity. Furthermore, the user must match the photograph on the identity document, so the mule must be present and cooperate with the attack. A stolen document cannot be used by an attacker, unless as a form of look-a-like fraud. The bigger problem is fraudulent identity documents; since they are hard to detect by the RA.

The remote and physical vetting processes differ in terms of how accessible they are to attackers. The current process requires physical presence at the RA desk. The remote vetting process can be exploited from anywhere in the world. Furthermore, since an attacker will typically look for an account that provides access to e.g. the whole network, other user accounts, etc, a successful attack on such an account can be followed by much more attack opportunities and attempts. Therefore the remote vetting process is easier to access for an attacker.

In order to mitigate the risks of remote vetting the following compensating control can be considered:

- Inform the user about the purchase of a second authentication factor via a separate, preferably validated channel, such as a validated email address (provided by the IDP, i.e. to send an email to), validated mobile phone number (provided by iDIN, i.e. to send an SMS to) or physical address (provided by iDIN, i.e. to send a letter to). Note that this control is also in place for means issuers in the eHerkenning trust framework.

This control could be added to the SURFnet levels of assurance framework.

3.6 conclusions and recommendations

Matching IDP identity assertions with assertions provided by iDIN, ReadID or IRMA is not trivial for several reasons.

First, the IDPs do not provide sufficient attributes to be able to uniquely match identity assertions with each other. Currently, only surname in combination with initials or first name can be used for matching purposes. This is not sufficient for achieving substantial assurance, i.e. the risk of false acceptances is too high. It is recommended that institutional IDPs start sharing the date of birth and (optionally) gender attributes. Asking IDPs to share even more information is challenging for various reasons: ability and willingness of institutions[23] to do so and because of privacy issues. Furthermore, it is recommended to make strict agreements with the institutions about the processing of source data by the IDPs. This is to 'clean up' the apparent lack of homogeneity of identity attribute values at the IDPs and to make matching more efficient. Rules to be enforced as a prerequisite for using SURFsecureID with remote vetting include the use of full first names (no initials),

[23] At least one institutional IDP already provides birth date attributes; another institution claims that technically it is quite simple to provision its IDP with birth date attributes.

data of birth and (optionally) gender in order to optimize the identity matching quality as much as possible. This may not be feasible at short term. During the course of the proof-of-concept it will become more clear how good the matching will be with the current personal data.

Second, given this lack of homogeneity across the values of the attributes provided, an extensive rule set is required to be able to resolve and match the attributes provided by the various providers. A rule set that is able to deal with a set of attributes such as surname, first names, initials and prefixes in various combinations and with variable values. Besides stricter rules, normalisation of identity information is required to simplify the matching process.

The recommended matching strategy is to match on full name and date of birth and optionally gender. Full name implies full first name(s) or initials, so rules are required that enable the matching algorithm to deal with this. In case this does not unambiguously lead to a match, extra measures can be taken. These (optional) measures are to ask the user to make use of a second external identity provider and use these asserted attributes to improve the matching or to ask the user for input. In case the attributes of the second identity provider do not match with those of the institutional IDP as well, the attributes will be matched with the attributes provided by the first external provider. Should these match, the user can continue; in case of a mismatch the RA will be involved. Furthermore, it is recommended to record the vetting means the user has used. This will simplify the matching process for subsequent new SURFsecureID authentication means requests as the user can be recommended to make use of the external identity provider that previously resulted in a successful identity match. It is also recommended to continuously monitor and evaluate the matching outcomes in order to refine this matching strategy. A fall back scenario should be in place for users that cannot be matched; they should be able to make use of the current physical process. From a privacy perspective: limit the storage duration of the identity attributes for matching purposes as short as possible, i.e. only for the duration of the vetting process.

Finally, in order to mitigate the extra risks of remote vetting – compared to physical vetting at the registration desk – a compensating measures should be considered: the ability to inform the user about the outcome of the vetting process via a separate and validated channel.

4 Functional Design

4.1 current process

For SURFsecureID, the registration and vetting processes to establish the identity of the user and to link this identity to his authentication credentials has been implemented as follows:

1. The user logs in at the SURFsecureID self-service registration portal with his federated institutional account. SURFsecureID receives an authentication assertion from the institutional identity provider that contains the first and last name of the user and his email address.
2. The user selects a second authentication factor (SMS, Tiqr, YubiKey, …) to be registered and does an authentication with it to prove that he owns the token. The token ID is linked to the user's institutional ID.
3. The user receives an e-mail and is asked to click on the activation link.
4. The user receives an activation code via e-mail.
5. The user goes to the registration authority (RA) and hands over the activation code.
6. The RA logs in into the SURFsecureID management portal and enters the activation code to find the corresponding token registration.
7. The RA asks the user to authenticate with the token to prove that he indeed owns the registered token.
8. The RA asks the user to show his identity document.
9. The RA checks the user's identity, i.e. compares the name of the user on the identity document with the name in the portal and compare the user's face with the picture on the identity document.
10. The RA enters the last 6 digits of the identity document number for accountability purposes.
11. The RA activates the token in the SURFsecureID management portal, i.e. the binding the between the user's verified identity, the user's identity at the IDP and his token is established. The user can now use the token as a second factor authentication credential.

Steps 1-4 constitute the self-service registration process; steps 5-9 constitute the identity vetting process. Step 10 is for audit/accountability purposes. These steps mimic the registration and physical vetting processes of e.g. authentication service providers in Idensys/eHerkenning such as Digidentity and KPN.

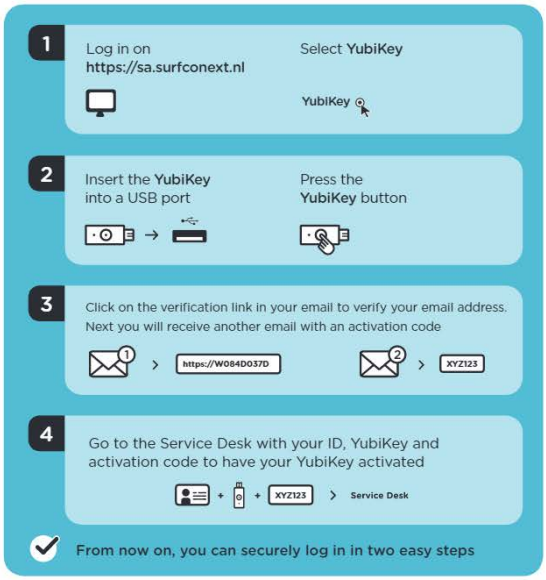


Figure 6: Part of the current process for obtaining a YubiKey token.

Each institute has one or a few RA Administrators (RAAs), who can assign the RA role to employees within their institute. RAs can only do the identity vetting for users from their own institute.

4.2 High-level flow and design decisions

On an abstract level, the current registration process with physical identity vetting process and the process with remote vetting consist of the steps visualised in Figure 7.

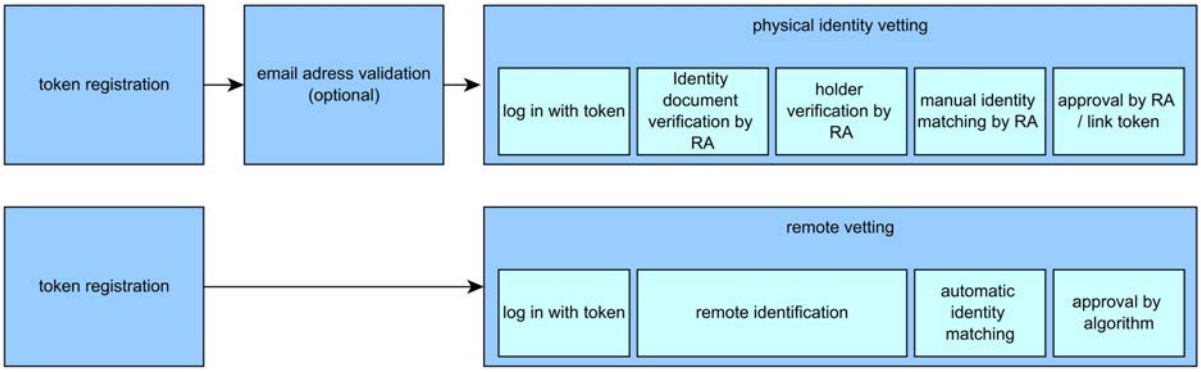


Figure 7: Current and remote registration processes

Below are detailed descriptions of the concrete steps needed to achieve the remote vetting. In designing these processes, we made the following overall design decisions:

1. The remote vetting process should by default stay as close as possible to the existing physical identity vetting process.
2. The email verification/activation is excluded from the remote vetting process. Although it may provide an additional 'what you know' factor into the process, the gained increase in the overall trust is small compared to the effort and potential issues. For instance, some institutions require 2nd factor authentication for access to email; thus creating a bootstrap paradox. Moreover, as a measure against man-in-the-browser attacks email verification is futile when an institution uses webmail, which most do. Finally, from a UX perspective email verification creates a hurdle for the user.
3. The user can choose between the three remote vetting options (in the PoC, for the final solution SURFnet can revisit this decision, e.g., the institute could choose).
4. The users have as much control as possible over the process, so that they don't unnecessarily get 'stuck'. This means they for instance can initiate a retry of remote identification themselves. This means that specifically users that fail the identification can restart the vetting at any time.
5. However, fraud by means of endless trial-and-error is prevented by restricting re-try's to a limited time period (e.g. 2 weeks) and limited amount of attempts (e.g. 5 attempts).
6. Users that have successfully completed the token registration cannot register more tokens of that type. However, if they fail to complete the subsequent remote identity vetting, their token registration expires after (for example) two weeks and can be redone. Registered tokens (including incompletely registered ones) can be removed by the user at any time.
7. Apart from the amount of attempts, there is no state kept in the remote identification process. If a retry is necessary, users can choose another remote identification method every time (thus allowing them to try another). Once the remote identification process is finished, they cannot do it again for that token registration.
8. The (automatic) identity matching will not be allowed to halt the process; i.e. "computer says no" will not be implemented. Instead, if the (automatic) identity matching fails, the vetting falls back to an RA. It should remain possible to for a human being to handle the vetting process, even if the user is remote. Preferably,

the RA of the institute of the user will handle these manual matches; but in case this isn't possible (or the user is not affiliated with any institute) there should be a meta or central RA.

9. Matching guidelines should be provided for the RA's, that are in line with guidelines for identity matching in the physical identity vetting process. Since there are no clear rules for what are acceptable matches in the physical process, such rules can be developed for both types of processes creating parity between them. The exact rules and instructions for manual matching are beyond the scope of this report.
10. For matches that fail automated processing, the RA gets digital / remote insight into the vetting process and the SURFconext and iDIN/ReadID/IRMA attributes, and then makes a decision; approve the matching, block the user, or initiate a retry. Blocking at least prevents the user from finishing the remote vetting process or initiating new remote vetting processes. Further details of what it means to block a user (including how this can be resolved) should be in line with SURFconext and SURFsecureID protocols and are out of scope for this report. The user is notified through the self-registration portal that the vetting is being processed by the RA.
11. Considering the design decisions above, it is necessary for the RA to be able to view anyone in the process and possibly intervene with a Retry or Block for everyone in there.
12. A record will be kept of which remote vetting method was used to obtain the second authentication factor. This can help in e.g. solving unhappy flows or renewing a token. It will also allow for possibly re-evaluating the LoA of the second factor.

4.3 High-level architecture view

Figure 8 depicts the high-level architecture.

Per remote vetting method the main interactions are:

- **iDIN** – the user is redirected by the SURFsecureID to the iDIN acquirer (or Digital Identity Service Provider), selects his bank, and is then redirected to that bank to log in. The actual login method verifies per bank, and if the user is using a mobile with a mobile banking app or not. After successful login and consent to the data exchange, this data is sent to the SURFsecureID.
- **ReadID** – the SURFsecureID server retrieves a (QR) token from the ReadID server, and shows this QR token to the user in a webpage that also contains an explanation on how to download the ReadID Ready app. The user installs the app and scans the QR code (or click on a link if the webpage is displayed on a mobile on which the ReadID Ready app is present). This links the ReadID session to the SSID token registration session. This binds the identity of the person to the token and the SURFsecureID account. After the user reads his chip, the chip data is sent to the ReadID server for verification, and the user takes a selfie that via the ReadID server is sent to a facial recognition provider (iProov), proving he is the rightful holder of the document This includes liveness checks. After the user is done, the overall results (chip + selfie) are retrieved by the SURFsecureID server.
- **IRMA/BRP** – we assume here that a user is not yet an IRMA user, and thus also still has to load the BRP attributes. The flow is similar to ReadID, in that the user has to install an app, in this case the IRMA app. The user then has to load the BRP attributes, which ones the user will be told by the SURFsecureID website. The user then selects in the IRMA app (or on the IRMA website) the option to load those, after which the user is directed to the Gemeente Nijmegen, and then to DigiD to log in with username/password + SMS (or DigiD App). The user accepts the loading. It can then scan a QR code provided by the SURFsecureID website and agrees to provide the needed attributes. This links the IRMA session to the SSID token registration session. This binds the identity of the person to the token and the SURFsecureID account.

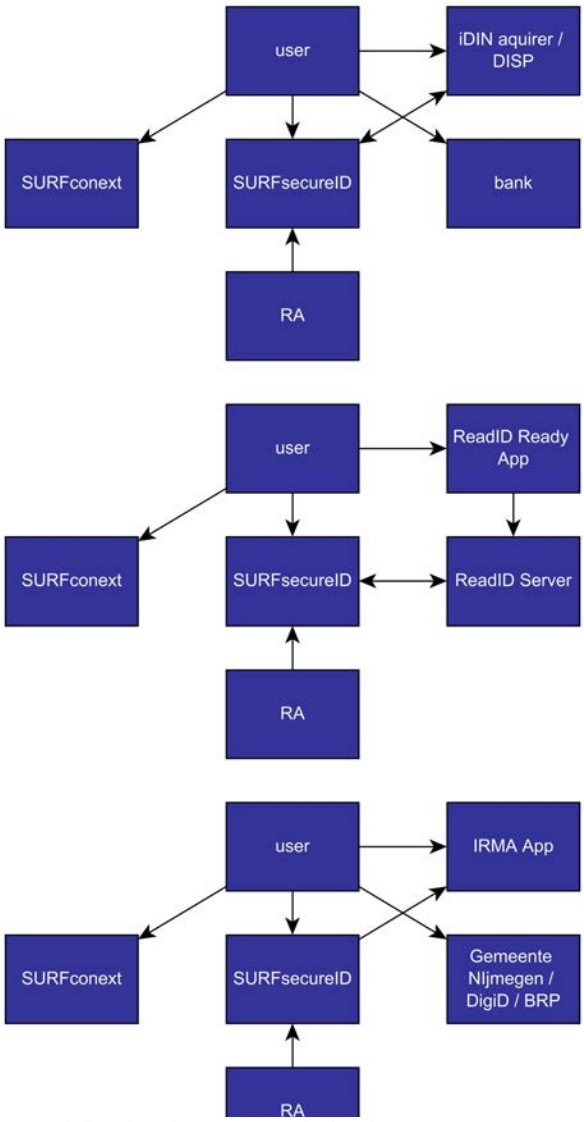


Figure 8: High-level architecture with the three remote vetting methods

4.4 Detailed flow

4.4.1 Overall flow

The remote vetting process is largely identical for each of the three identification options; the details of these are discussed below. See Figure 9 for a visual overview of the remote vetting process.

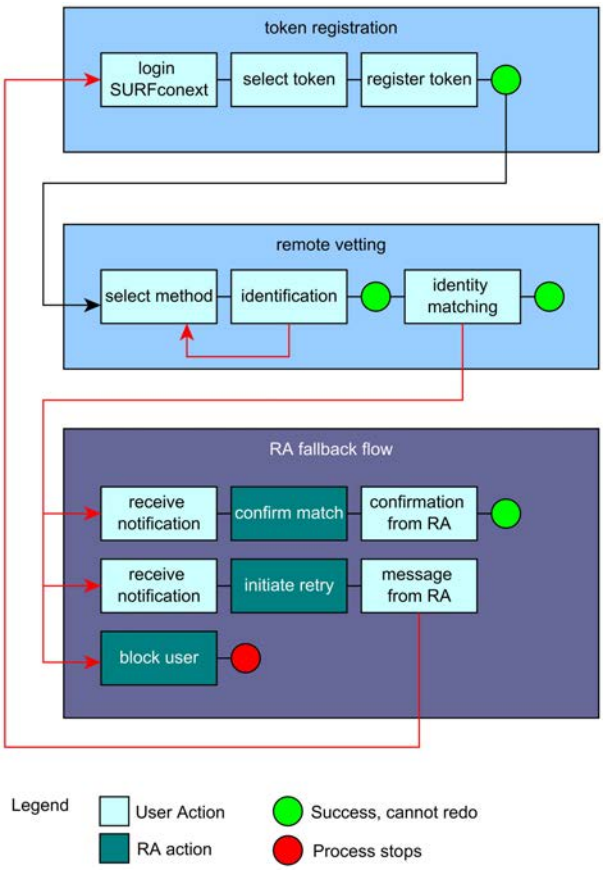


Figure 9: detailed steps in the remote vetting process

1. The user logs in at the SURFsecureID self-service registration portal with his federated institution account. SURFsecureID receives an authentication assertion from the institutional identity provider that contains the first and last name, date of birth and (optionally) gender of the user.
2. The user selects a second authentication factor (SMS, Tiqr, YubiKey, …) to register, and performs an authentication with it to prove that he owns the factor.
3. The user arrives on a webpage where the identification methods are explained and can be selected. The user selects one.
4. The user executes the identification steps (see sections 4.4.3, 4.4.4 and 4.4.5 below). When the identification is completed successfully, this cannot be redone for that token.
5. In the backend, the identity (attributes) provided from the identification are matched with the SURFconext identity. When a match is successfully established, the token is activated and linked to the SURFsecureID (and this SURFconext) account.
6. In case the identity matching fails, the user can retry. If this again fails the process is handed over to the RA. This breaks the straight-through processing. The user receives a notification that the process is being handled by the RA, via the web interface. In case an institute does not have an RA, a central RA (e.g., from SURFnet) needs to do this step (as is the case in the current physical registration process).
7. The RA makes a decision. There are three possibilities:
    a. The RA manually approves the match remotely, with the same instructions for this as the face-to-face process has if the attributes do not 100% match.
    b. The RA initiates a retry for the user, resetting the whole process
    c. The RA blocks the SURFconext account (in case of fraud suspicion, the user cannot retry / add a new token without the RA unblocking the account).
8. The user either
    a. Is notified of the approved match and finished process. No further actions are required.
    b. Is notified to retry the whole registration process. The RA can include a personalized message to explain the situation and/or advising on how to retry.
    c. Is blocked and does not receive any notification of this, to prevent providing fraudsters with too much information.

4.4.2 Remote identification flows

The remote vetting uses pre-existing identification solutions, that means each identification method has an already established user flow. These are summarised in Figure 10.

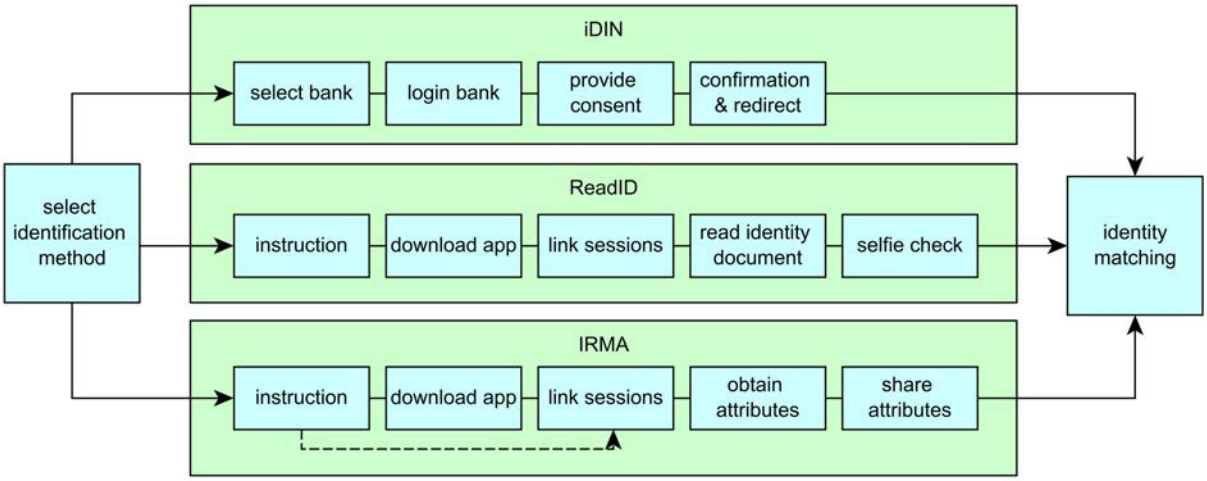


Figure 10: iDIN, ReadID and IRMA identification flows

4.4.3 iDIN details

The iDIN identification process is as follows:

1. the user selects the bank to be used for the iDIN identification.
2. The user logs in with his credentials for that bank. How this is done of course differs per bank.
3. The identity attributes to be shared are shown: last name, initials, date of birth dan gender. The user provides his consent to share these attributes.
4. The reception of the attributes is confirmed by iDIN, and the user is redirected back to the SURFsecureID self-service portal.
5. Meanwhile, in the backend, the attributes are communicated with SURFsecureID.
6. Identity matching is performed by SURFsecureID.

4.4.4 ReadID details

The ReadID identification process consists of the following steps:

1. The user receives an instruction to download the ReadID Ready application. Or in case it's already installed, to simply open it. ReadID Ready is a ready-to-use app, so that SURFnet does not have to implement an own app on top of the ReadID SDK.
2. The user downloads the app.
3. The app is linked to the user's SURFconext account, either via scanning a QR code shown in the SURFsecureID portal or by clicking an activation link (the latter if the user is using a mobile device and the ReadID Ready app is installed on this same device).
4. The user reads his legal identity document with the ReadID Ready app.
5. The user performs a selfie-check to confirm he is the rightful holder of the identity document. This decreases the risk of identity fraud.
6. The outcomes of the identity document verification are communicated with SURFsecureID.
7. Identity matching is performed by SURFsecureID.

4.4.5 IRMA details

The IRMA identification process encompasses the following steps:

1. The user receives an instruction to download the IRMA app.
2. The user downloads the app and creates an IRMA account. If the user previously has downloaded IRMA, step 2 can be skipped.

3. The user obtains the identity attributes from the BRP. These attributes have to be fresh to increase trust, i.e., they cannot be cached from e.g. 60 days ago. The user thus has to do the below in all cases:
    a. The user is directed to the website of Gemeente Nijmegen, where he logs in with DigiD Midden.
    b. The user consents to importing the attributes into IRMA.
    c. The user is directed back to IRMA.
4. The user shares the BRP identity attributes – full name, date of birth and gender - with SURFsecureID. This is done by scanning a QR code that is displayed in the SURFsecureID portal.
5. Identity matching is performed by SURFsecureID.

Optional: force users to do also do iDIN, to increase trust level.

5 Level of Assurance Analysis

5.1 Risk factors and controls analysis

A number of risks can be identified that may need mitigating controls. Under the assumption that the first factor has been compromised (i.e. the user's institutional account) these risks are:

- A fraudulent/fake identity document is used.
- A stolen identity document is used.
- Someone with the same name as the rightful holder of the first factor tries to obtain a second factor.
- There is a man in the middle that manipulates the communication in order to get access to the second factor, i.e. the user is unknowingly registering the second factor of the attacker.
- Social engineering, i.e. someone is being tricked into performing the remote vetting process in a real-life interaction with another person, e.g. someone pretending to be a postman comes at the door and requires "authentication" in order to receive a valuable package.
- Phishing, i.e. someone is tricked into performing the remote vetting process by fully digital means, e.g. someone receives an email with URL from their "boss" telling them they need to authenticate.
- The smartphone is compromised; someone has physical access to the rightful person's smartphone.
- Evil twin / look-a-like fraud; someone tries to impersonate a rightful person they physically resemble or they resemble in name.
- Trial and error / guessing, someone just tries to get a positive identity match by logging in with iDIN/ReadID/IRMA multiple times in a row.
- iDIN or IRMA itself is compromised; the attacker has access to the credentials needed to use these authentication means. ReadID does not involve credentials.

The table below provides an overview of risk mitigating measures in the current situation and for the future iDIN, ReadID and IRMA solutions. The scope of the analysis below is limited to the vetting process only, not to SURFconext or SURFsecureID authentication in general.

Risk	Current vetting process	iDIN	ReadID	IRMA
Fake/fraudulent identity document	Trained RA and tool-support (currently not done)	Not applicable; identity document verification is part of the iDIN issuing process executed by banks	Check authenticity of digital signature of the attributes obtained.	Not applicable
Stolen identity document	RA compares user with face image on the document.	Not applicable; identity document verification is part of the iDIN issuing process executed by banks	Selfie-check that compares selfie with facial image read from the chip. Optionally, a check for lost/stolen documents at VIS.	Not applicable

Same name	Check for date of birth and possible other attributes (currently not part of the process)	Check for date of birth and possible other attributes	Check for date of birth and possible other attributes	Check for date of birth and possible other attributes
Man in the middle	Provide TLS-encryption. Use of out-of-band channels for communication (e.g. use of email or activation code in combination with 2nd factor authentication to mitigate man in the browser attacks).	Provide TLS-encryption	Provide TLS-encryption	Provide TLS-encryption
Social engineering	User awareness and visibility of activity	User awareness and visibility of activity	User awareness and visibility of activity	User awareness and visibility of activity
Phishing	User awareness	User awareness	User awareness	User awareness
Compromised phone	Tiqr requires a fingerprint or PIN;. SMS doesn't have such additional security.	Mobile banking apps have PIN-codes to prevent unauthorised access	Not applicable	IRMA-app has PIN-code to prevent unauthorised use for remote vetting. PIN is checked remotely, so compromise of mobile phone does not give PIN.
Look-a-like fraud	Trained RA, but difficult to mitigate.	Requires access to iDIN credentials as well	Selfie check based on the high-resolution facial image from the document chip	Requires access to IRMA credentials as well

Trial and error	RA will notice that the same person tries to get a second factor multiple times. This however is not a strong control; RA has no log of rejected vetting attempts, nor clear instructions.	Limit the number of iDIN attempts (and remote vetting attempts in general)	Limit the number of ReadID attempt (and remote vetting attempts in general)	Limit the number of IRMA attempts (and remote vetting attempts in general)
iDIN or IRMA is compromised	Not applicable	User awareness and visibility of activity	Not applicable; for equivalent see stolen and fake identity document and look-a-like fraud	Enforcing a real-time loading of the attributes from the BRP introduces another authentication means (DigiD) to which the attacker must have access.

There will always be users that cannot be vetted remotely. For these users fallback solutions must be in place, i.e. the current physical identification process.

5.2 Level of Assurance SURFsecureID

Obviously, these risks have an impact on the overall authentication assurance level of SURFsecureID. SURFnet has adopted its own level of assurance (LoA) framework that is largely based on ISO29115.[24] The resulting LoA of the current physical process depends on the authentication factor and takes a number of enrolment requirements into account:

- LoA 1: Password authentication through SURFconext at the user's home IdP;
- LoA 2: LoA 1 + SMS or Tiqr authentication;
- LoA 3: LoA 1 + YubiKey (hardware token) authentication.
- LoA 4: Not available.

It is assumed that proper measures are taken to prevent authentication protocol threats such as eavesdropping, man-in-the-middle, replaying, and hijacking. Attacks are not limited to the authentication protocol itself. Other attacks include the use of malicious code to compromise authentication tokens, insider threats to compromise authentication tokens, social engineering to get a subscriber to reveal his password to the attacker, "shoulder-surfing", fooling claimants into using an insecure protocol, when they think that they are using a secure protocol, or intentionally denying ever having registered by subscribers who deliberately compromise their tokens.

[24] SURFsecureID Levels of Assurance, see `https://wiki.surfnet.nl/display/SsID/Levels+of+Assurance`.

The ReadID-with-selfie remote vetting is the most literal digital version of the existing process, with the assessment of validity of the identity document as well of the holder verification being replaced with software. This is not susceptible to social engineering and facial recognition by modern software is as good as by humans[25]. A sufficiently rich set of attributes is obtained to cater for optimal matching with the IDP-provided attributes. LoA 3 is achievable with ReadID, possibly LoA 4.

The LoA of iDIN is eIDAS Substantial compliant. However the details of the issuing process (and the process of opening a bank account) differ per bank. Several Dutch banks use the same ReadID-with-selfie process as in the remote vetting PoC. But users may have gotten it through other means, including processes similar to the current physical vetting process. All Dutch banks have to comply to the AML directive and are subject to supervision by DNB, and iDIN is positioned as a possible private alternative for DigiD Substantial, so there seems to be a consensus that iDIN is roughly eIDAS substantial level. In the context of the SURFnet LoA framework this translates to LoA 3. Specific point to note here is that it is common for banks to rely on a derived identity from another bank, thus getting a bank account based on another bank account resulting in a very long trust chain. DNB and other authorities are expected to provide guidance that this is no longer allowed. Using iDIN for SURFsecureID has a similar disadvantage, but from a trust perspective iDIN is likely good enough anyway for SURFsecureID.

IRMA has a lower LoA compared to the other two remote vetting solutions and the current process, mainly because of the usage of DigiD Midden (which is eIDAS Low or LoA 2 in terms of ISO29115). But this could change, by using DigiD Substantial (which is comparable with eIDAS Substantial or ISO29115 LoA 3).

5.3 The SURFsecureID level of assurance framework

As said, the SURFnet LoA framework focusses in particular on the strength of the authentication factor. Little has been formalised regarding the level of identity assurance, or in other words, the quality of the registration and enrolment process. The nuance here is that RAs are generally trusted and the institutes are the primary users of SURFsecureID, and thus have an inherent interest into doing this properly. However, with remote vetting formalizing the identity assurance process does become more important. Moreover, also the assurance quality of the identity matching should be taken into account as part of the identity assurance. This factor, however, is not easily translated to authentication assurance levels. Only the NIST framework gives some guidance in the respect; other frameworks such as eIDAS or ISO29115 do not address this matter. Only after the evaluation of the proof-of-concept it is possible to determine the quality of the proposed matching strategy and take this into account for the determination of the LoA of the remote vetting solutions.

Based on the LoA frameworks of eIDAS and eHerkenning the following requirements are recommended for SURFnet to include in its own LoA framework for levels 2 and 3:

- Ensure the user is aware of recommended security precautions related to the SURFsecureID authentication means.
- (from section 3.5.3) Inform the user about the purchase of a second authentication factor via a separate, preferably validated channel, such as email address (provided by the IDP, i.e. to send an email to), mobile phone number (provided by IRMA, i.e. to send an SMS to) or physical address (provided by iDIN, i.e. to send a letter to).
- Define the levels of assurance for iDIN, ReadID and IRMA. Suggestion:
  - iDIN = LoA 3
  - ReadID = LoA 3

[25] Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms, 2017, see https://www.pnas.org/content/115/24/6171.

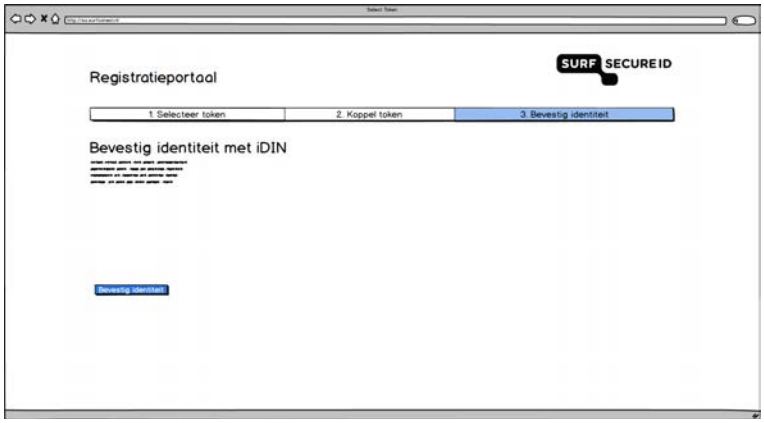
- IRMA = LoA2 (depending on the assurance level of DigiD used)

In conclusion, while iDIN and ReadID deliver a substantial second factor authentication means, IRMA may not. It would be fruitful to consider assigning several levels of assurance; each remote vetting means has its own level, rather than trying to unify the three remote vetting processes into one LoA. This mirrors the differing LoA levels for the hardware tokens. However, since the combined LoA's of token and remote vetting process are determined by the lowest LoA, in some cases the advantages of the higher LoA of iDIN and ReadID may be lost. One such LoA decreasing factor is the quality of the identity matching strategy. Poor matching outcomes (i.e. with many false acceptances) reduce the overall assurance level of the user's identity. Due to lack of matching experiences, it is currently hard to assess the LoA of SURFsecureID. Only after the evaluation of the proof-of-concept we may be able to conduct such an assessment.

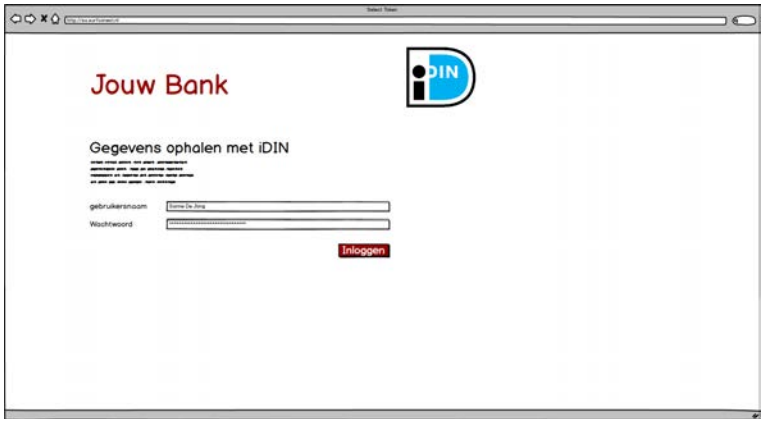
6 Mock-ups

Below are screen mock-ups of the three remote vetting flows. Please see the separate PDF & Balsamiq file for high resolution, zoomable versions. The first part of the remote vetting process, which consist of registering a second factor token, remains unchanged. Mock-ups of this part of the flow can be found in appendix A

Screen: choose identification method
User action: choose preferred method for remote identification
Status: to be built from scratch
Note: on this screen, the flows converge, and from this screen onwards, flows diverge again.



Screen: choose iDIN (iDIN flow)
User action: read instruction and confirm choice for iDIN
Status: to be built from scratch

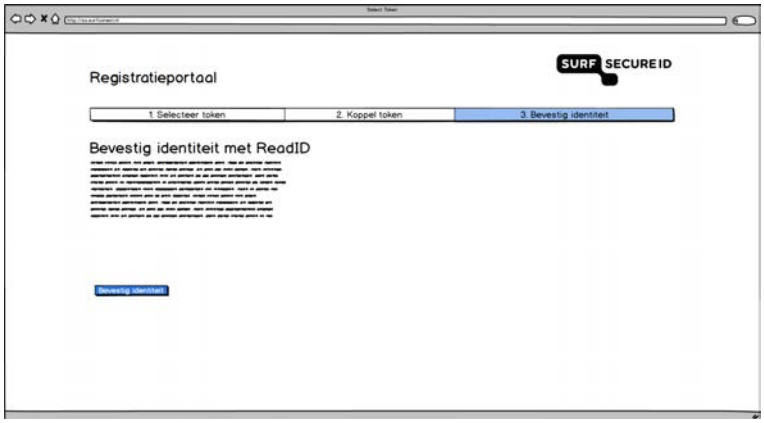


Screen: login to bank (iDIN flow)

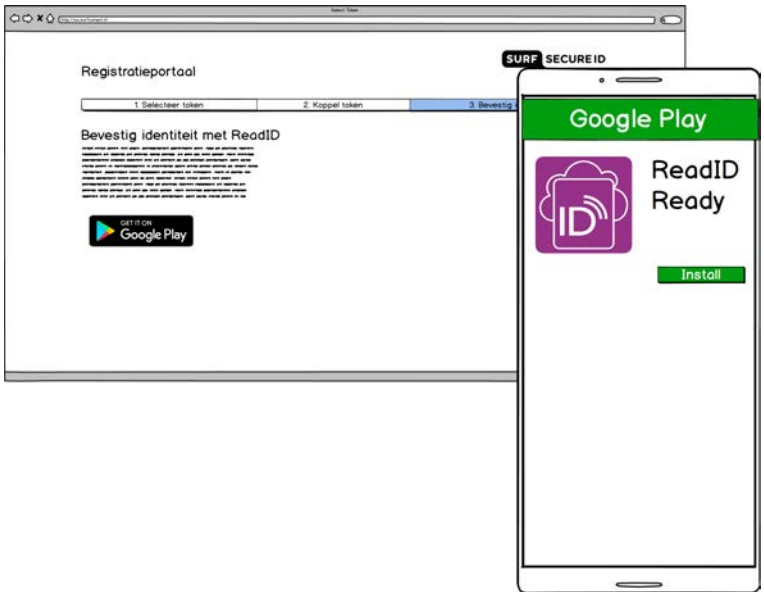
User action: get phone ready
Status: external, i.e. existing and ready to use, but not under the control of SURFnet

Screen: provide consent (iDIN flow)
User action: read text, check data and press confirm button
Status: external, i.e. existing and ready to use, but not under the control of SURFnet

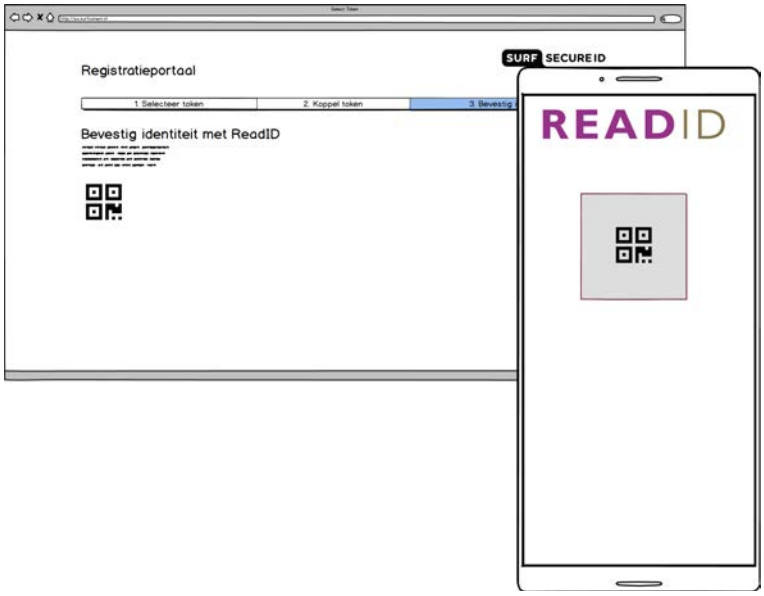
Screen: confirmation and re-direct (iDIN flow)
User action: press confirm button
Status: external, i.e. existing and ready to use, but not under the control of SURFnet



Screen: choose ReadID (ReadID flow)
User action: read instruction, confirm choosing ReadID
Status: to be built from scratch

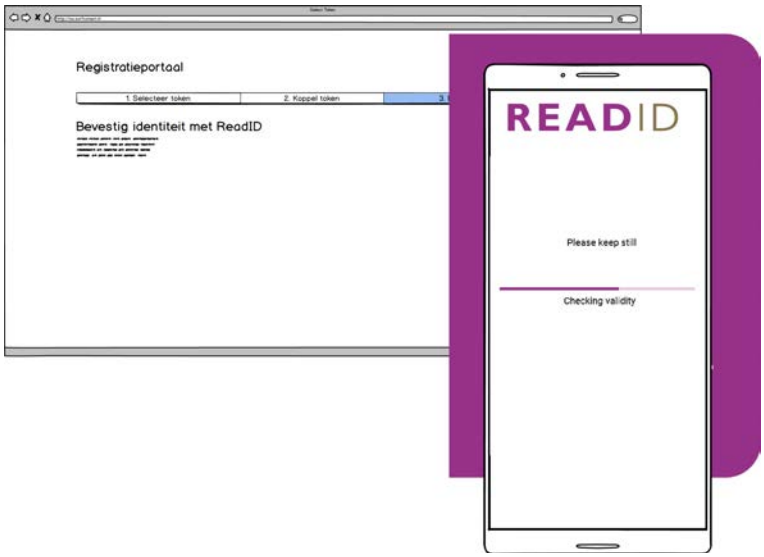


Screen: download ReadID Ready app (ReadID flow)
User action: go to Google play, download and install app
Status website: to be built from scratch
Status App: external, i.e. existing and ready to use, but not under the control of SURFnet

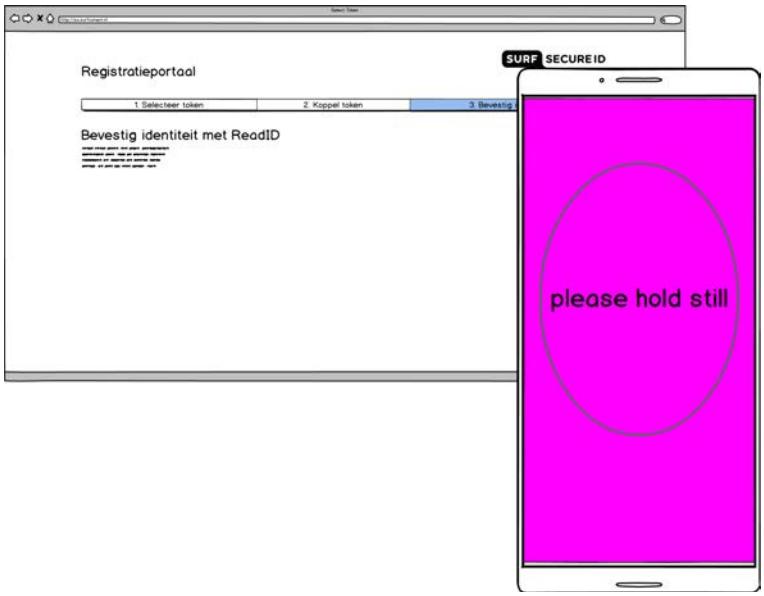


Screen: scan QR code to link sessions (ReadID flow)
User action: scan QR code on website with ReadID Ready app
Status website: to be built from scratch
Status App: external, i.e. existing and ready to use, but not under the control of SURFnet

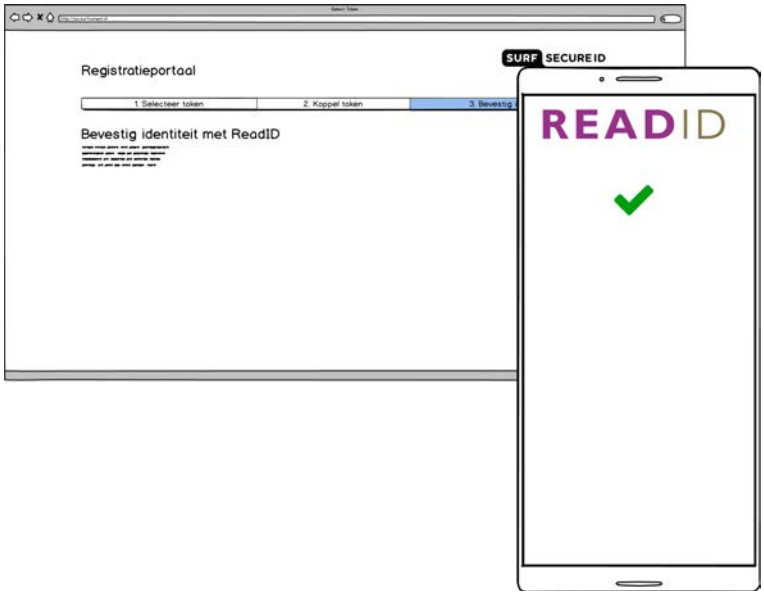
Screen: scan identity document (ReadID flow)
User action: scan MRZ on the identity document with ReadID Ready app
Status website: to be built from scratch
Status App: external, i.e. existing and ready to use, but not under the control of SURFnet



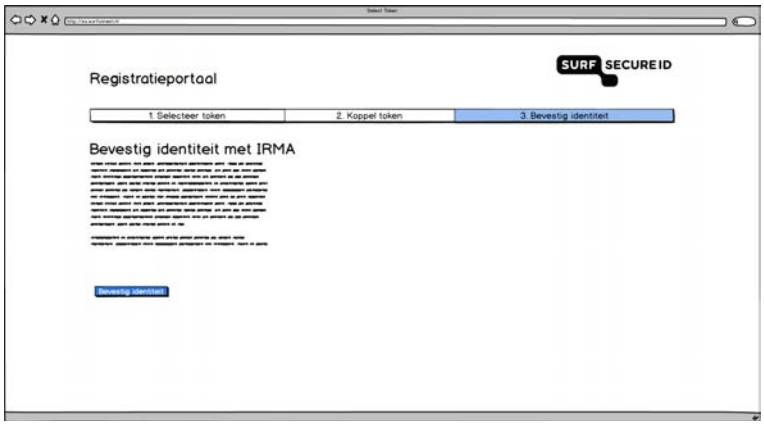
Screen: read chip of identity document (ReadID flow)
User action: read the chip of the identity document with ReadID Ready app
Status website: to be built from scratch
Status App: external, i.e. existing and ready to use, but not under the control of SURFnet



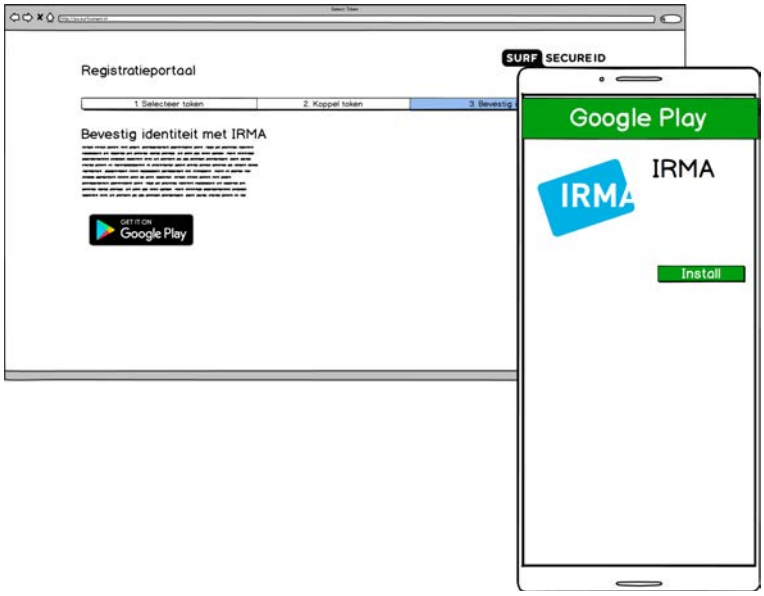
Screen: facial matching (ReadID flow)
User action: make a special selfie with the ReadID Ready app
Status website: to be built from scratch
Status App: external, i.e. existing and ready to use, but not under the control of SURFnet



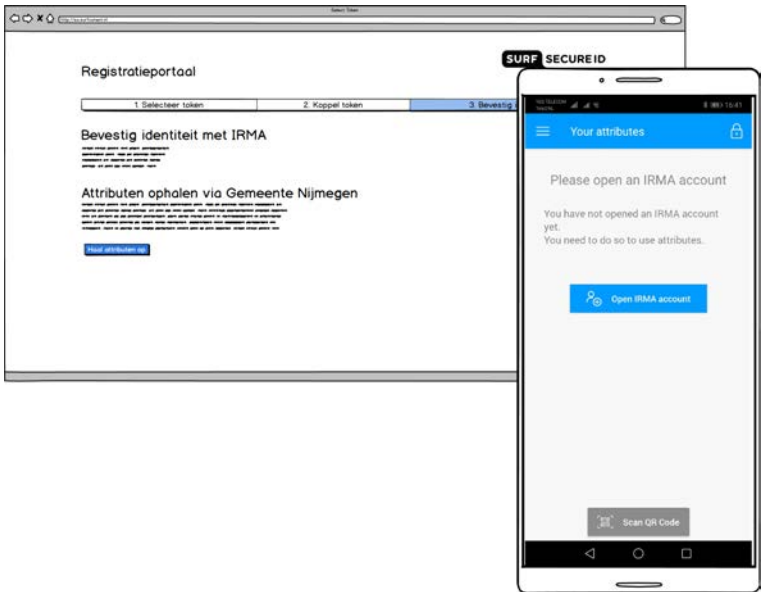
Screen: confirmation (ReadID flow)
User action: close app
Status website: to be built from scratch
Status App: external, i.e. existing and ready to use, but not under the control of SURFnet



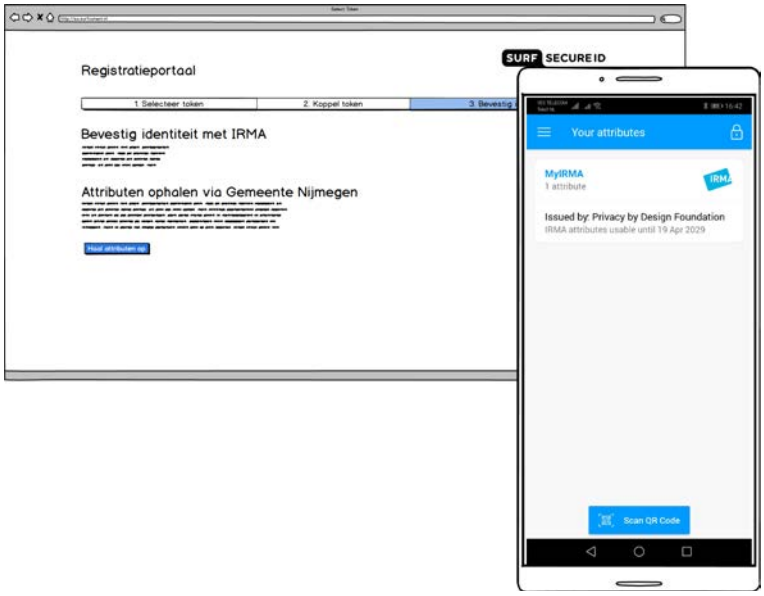
Screen: choose IRMA (IRMA flow)
User action: read instruction, confirm choosing IRMA
Status: to be built from scratch



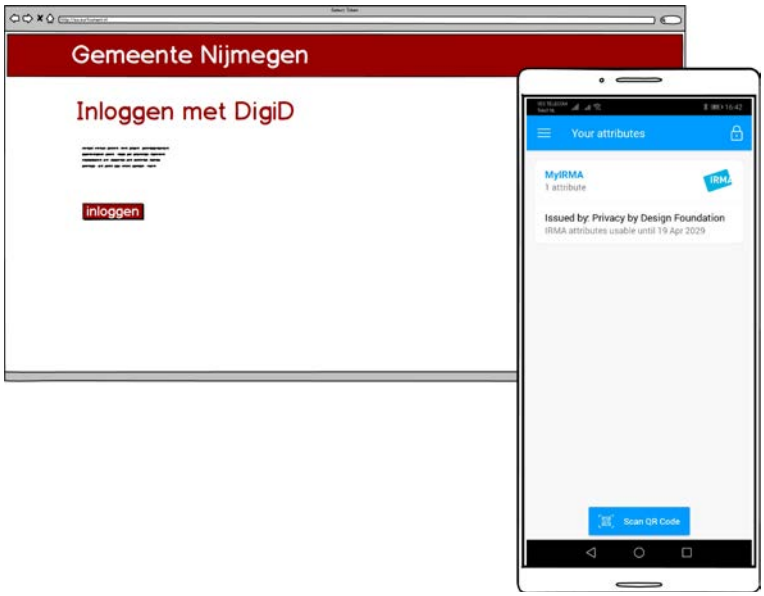
Screen: download IRMA app (IRMA flow)
User action: go to Google play and download and install IRMA app
Status website: to be built from scratch
Status app: external, i.e. existing and ready to use, but not under the control of SURFnet



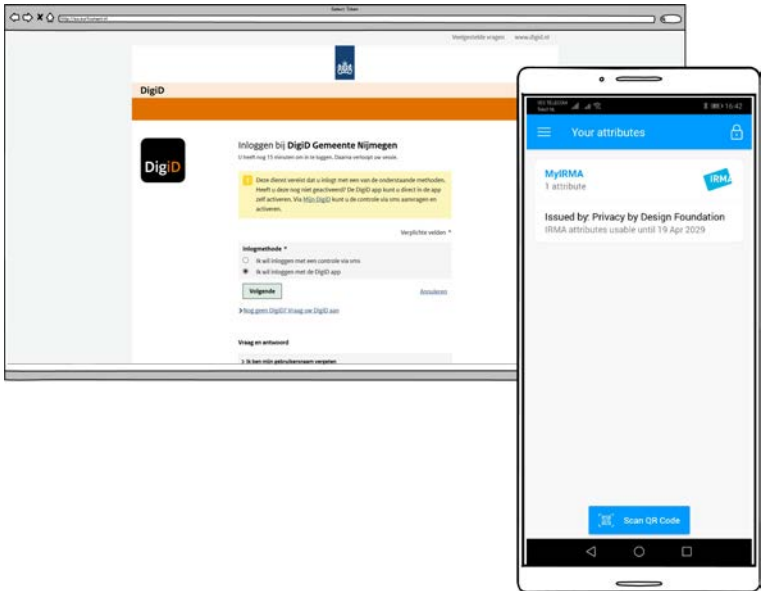
Screen: open IRMA account (IRMA flow)
User action: read instructions and follow steps to open an IRMA account
Status website: to be built from scratch
Status app: external, i.e. existing and ready to use, but not under the control of SURFnet



Screen: start obtaining attributes from Gemeente Nijmegen (IRMA flow)
User action: read instructions and confirm
Status website: to be built from scratch
Status app: external, i.e. existing and ready to use, but not under the control of SURFnet

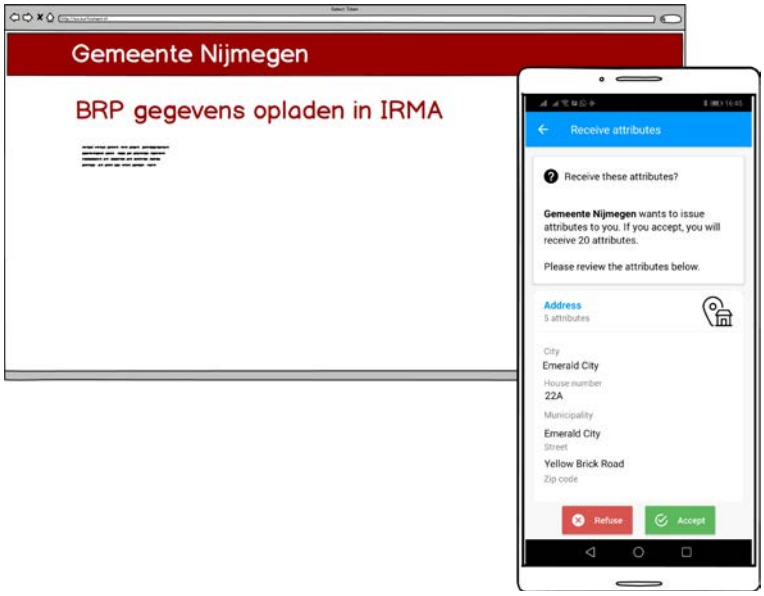


Screen: instruction Gemeente Nijmegen (IRMA flow)
User action: read instructions and confirm
Status website: external, i.e. existing and ready to use, but not under the control of SURFnet
Status app: external, i.e. existing and ready to use, but not under the control of SURFnet

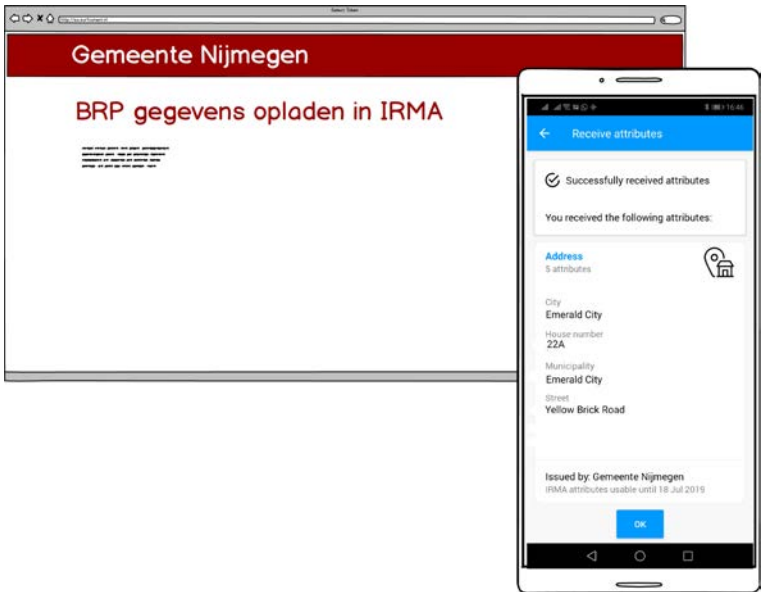


Screen: log in with DigiD to Gemeente Nijmegen (IRMA flow)
User action: enter DigiD credentials and log in
Status website: external, i.e. existing and ready to use, but not under the control of SURFnet
Status app: external, i.e. existing and ready to use, but not under the control of SURFnet

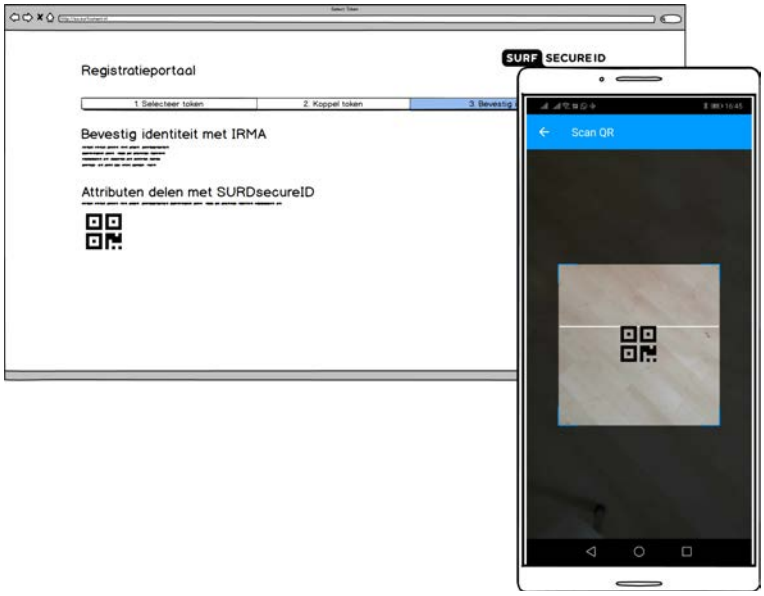
Screen: scan QR to obtain attributes from Gemeente Nijmegen (IRMA flow)
User action: scan QR with IRMA app
Status website: external, i.e. existing and ready to use, but not under the control of SURFnet
Status app: external, i.e. existing and ready to use, but not under the control of SURFnet



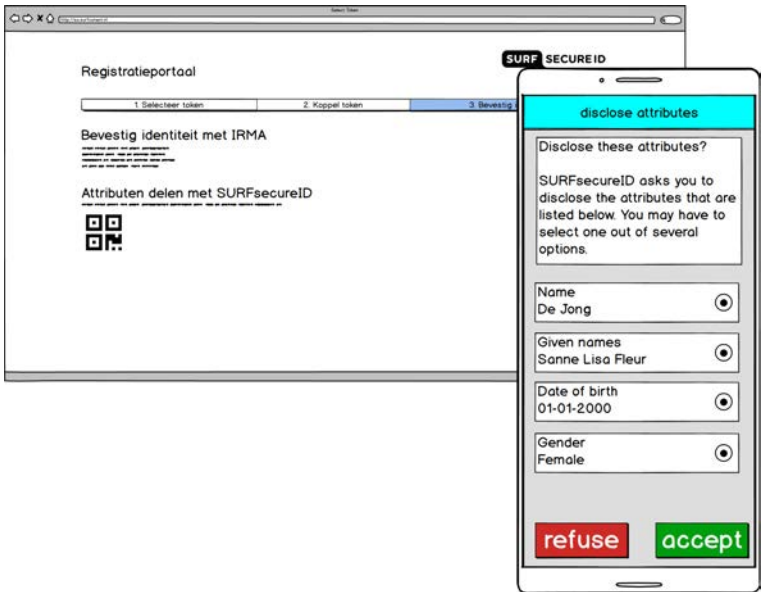
Screen: accept attributes from Gemeente Nijmegen (IRMA flow)
User action: read instructions and confirm
Status website: external, i.e. existing and ready to use, but not under the control of SURFnet
Status app: external, i.e. existing and ready to use, but not under the control of SURFnet



Screen: confirm reception of attributes from Gemeente Nijmegen (IRMA flow)
User action: read instructions and confirm
Status website: external, i.e. existing and ready to use, but not under the control of SURFnet
Status app: external, i.e. existing and ready to use, but not under the control of SURFnet



Screen: scan QR to link sessions and share attributes (IRMA flow)
User action: scan the QR code on the website with the IRMA app
Status website: to be built from scratch
Status app: external, i.e. existing and ready to use, but not under the control of SURFnet



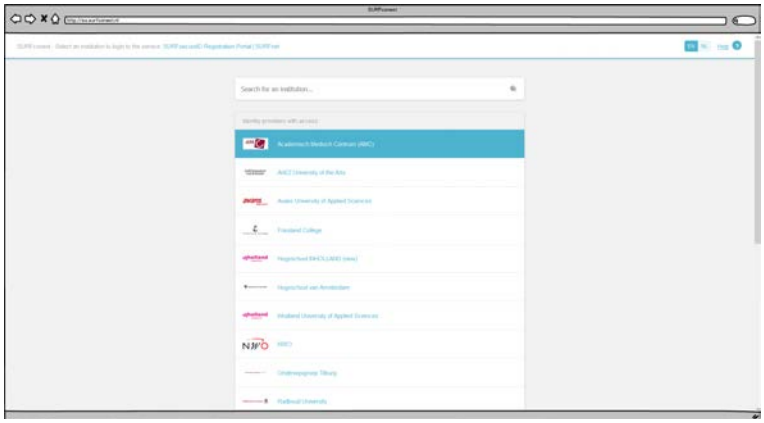
Screen: provide consent to share attributes (IRMA flow)
User action: read instructions, review attributes and provide consent
Status website: to be built from scratch
Status app: external, i.e. existing and ready to use, but not under the control of SURFnet

Screen: reception of data/attributes and identity matching
User action: wait for a few moments
Backend action: match received identity form remote identification with institutional identity
Status: to be built from scratch
Note: all flows converge on this screen

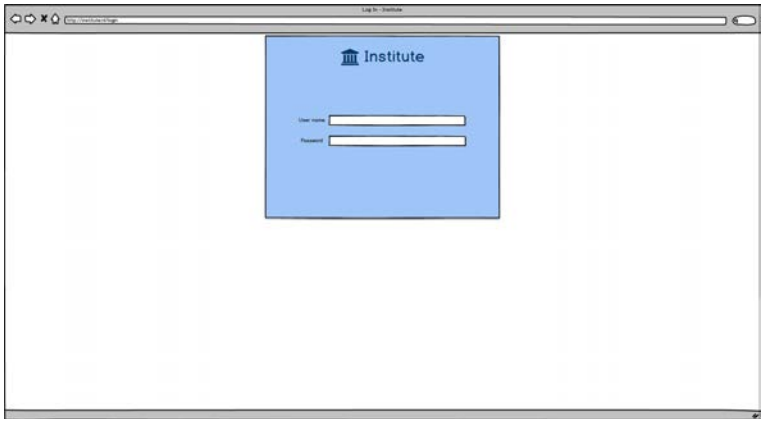
Screen: confirmation of successful registration
User action: be happy and close the website
Status: to be adapted from currently existing screens

Screen: notification of unsuccessful match
User action: check attributes and retry the remote vetting with another identification method
Status: to be adapted from currently existing screens
Note: if the user tries a second remote vetting method, the process starts gain from the first screen. The flow for the user is the same. In the backend however the results from the first remote vetting method are kept.

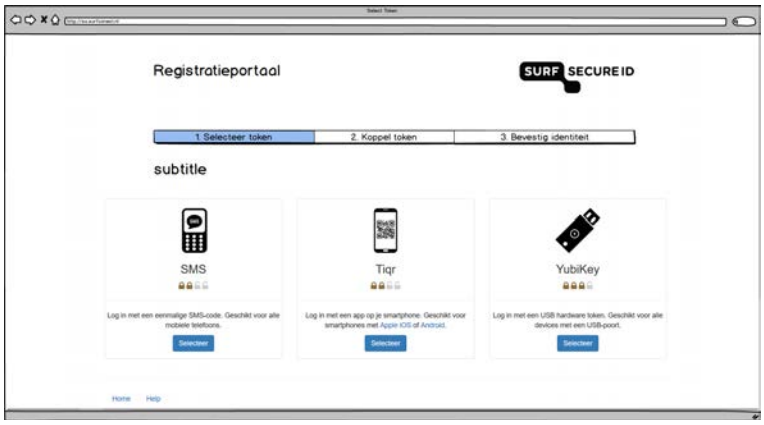
Appendix A: Mockups token registration



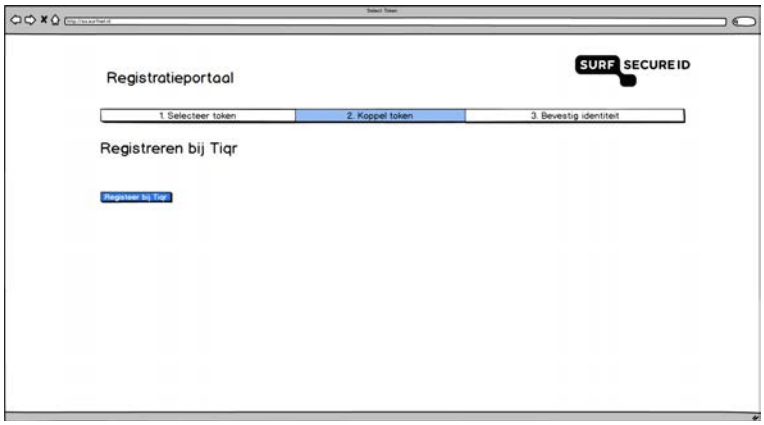
Screen: Select institute
User action: select his own institute for institutional login
Status: already existing, without adaptation necessary



Screen: institutional login
User action: fill in institute credentials
Status: external, i.e. existing and ready to use, but not under the control of SURFnet

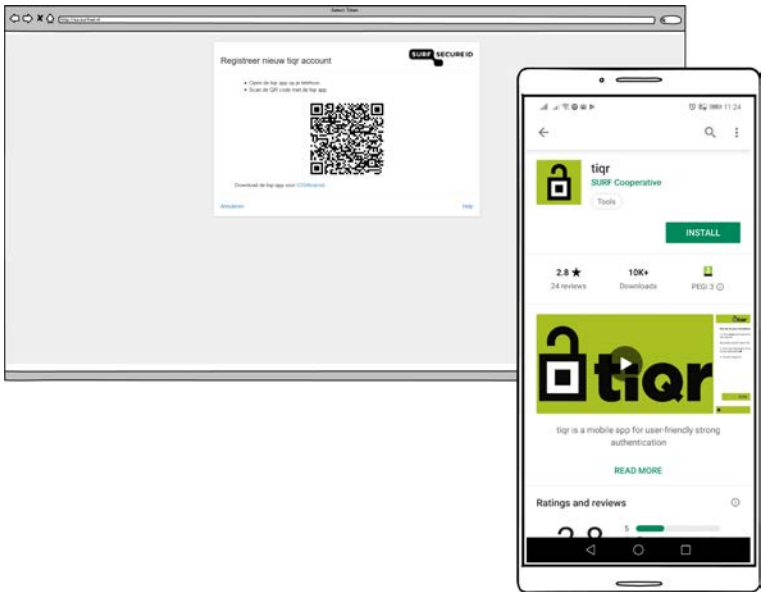


Screen: choose token
User action: choose the desired token to be registered, either SMS, Tiqr app, or YubiKey
Status: already existing, without adaptation necessary
Note: from this screen onwards, flows diverge

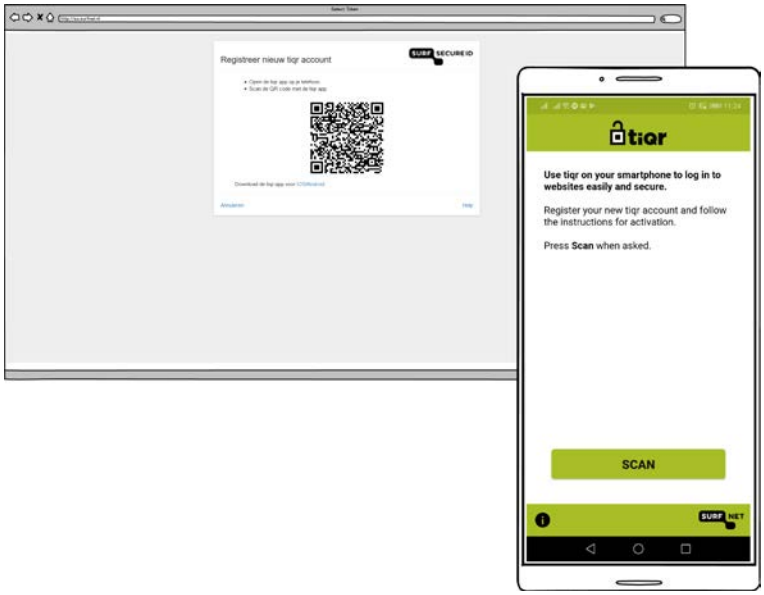


Screen: Tiqr registration (Tiqr Flow)
User action: start Tiqr registration
Status: already existing, without adaptation necessary

Screen: Tiqr registration QR (Tiqr Flow)
User action: get phone ready
Status: already existing, without adaptation necessary

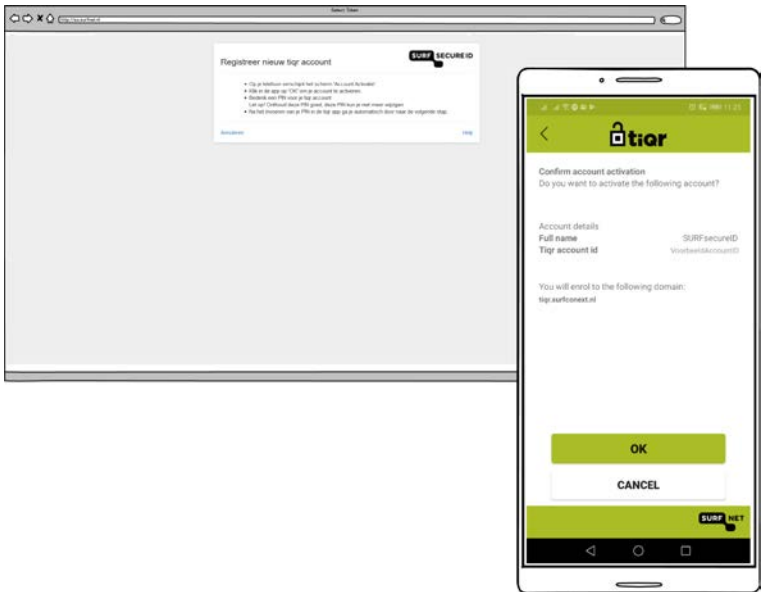


Screen: instal Tiqr (Tiqr Flow)
User action: download Tiqr app from Google play and install the app
Status: already existing, without adaptation necessary

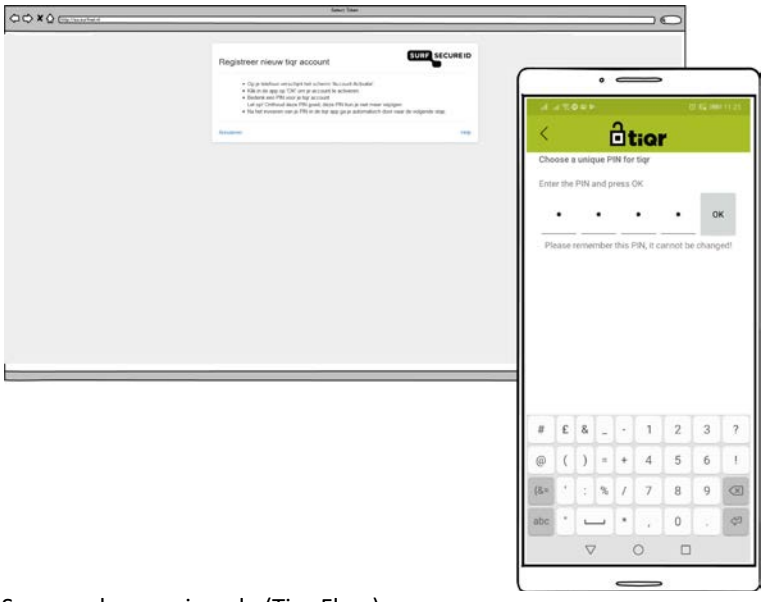


Screen: Tiqr scan instruction (Tiqr Flow)
User action: read instruction
Status: already existing, without adaptation necessary

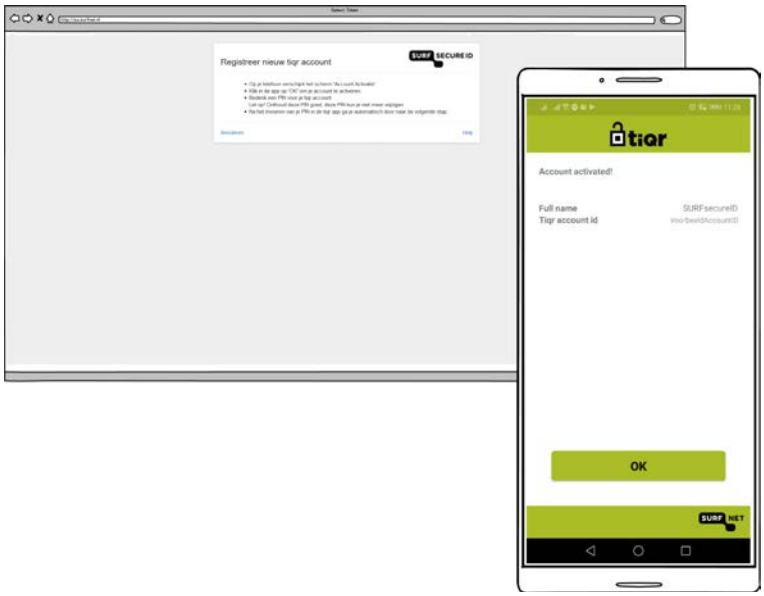
Screen: Scan QR code (Tiqr Flow)
User action: scan QR code on website with phone
Status: already existing, without adaptation necessary



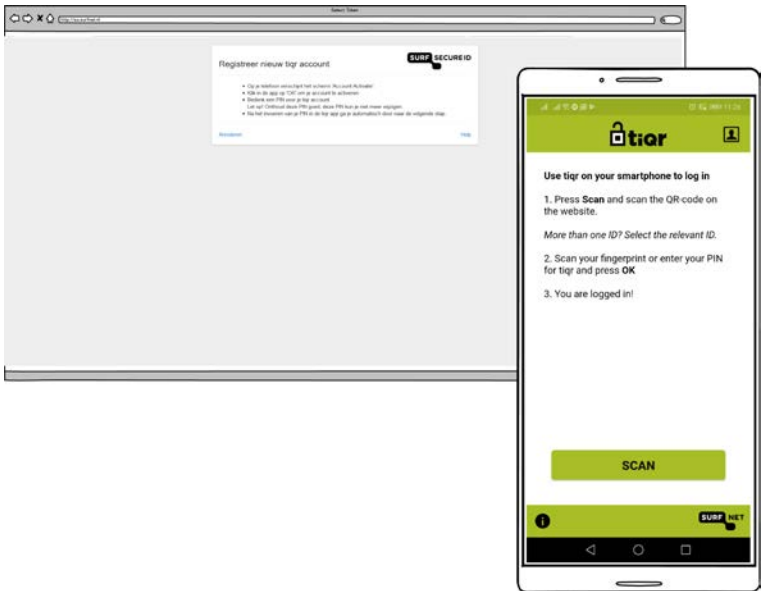
Screen: Tiqr confirm account activation (Tiqr Flow)
User action: read details and confirm
Status: already existing, without adaptation necessary



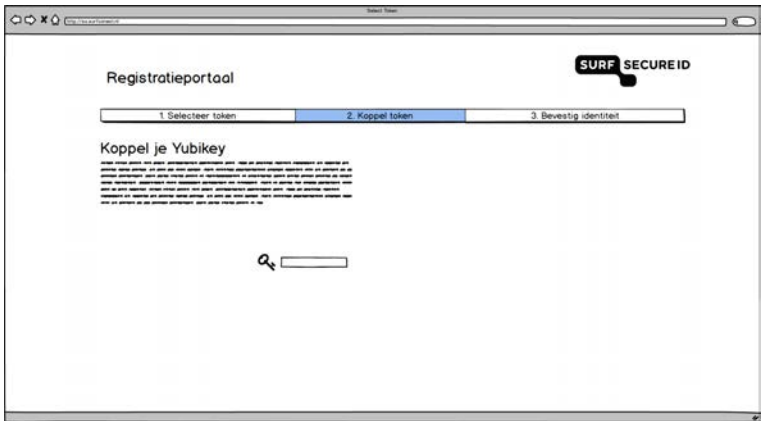
Screen: choose pincode (Tiqr Flow)
User action: choose a pincode
Status: already existing, without adaptation necessary



Screen: Tiqr confirmation (Tiqr Flow)
User action: read confirmation and press ok
Status: already existing, without adaptation necessary



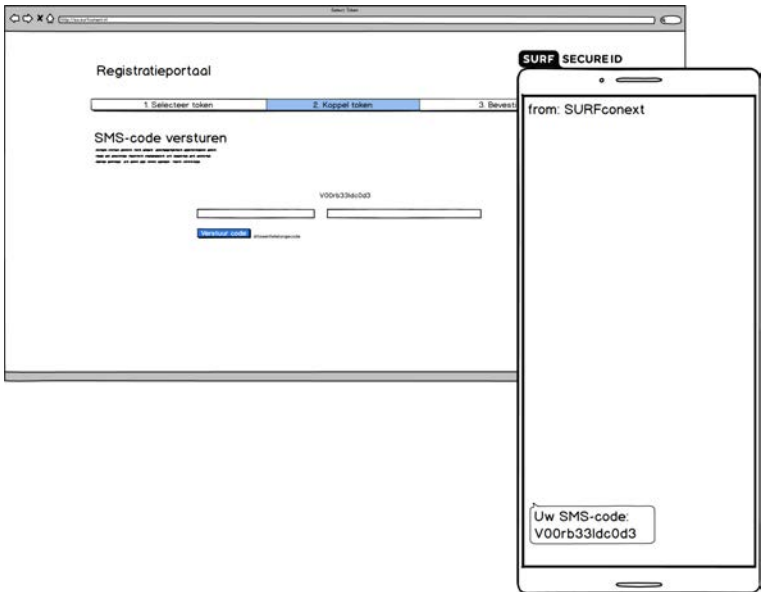
Screen: additional instructions on how to use Tiqr (Tiqr Flow)
User action: read text
Status: already existing, without adaptation necessary



Screen: Yubikey registration (YubiKey Flow)
User action: put Yubikey into USB port
Status: already existing, without adaptation necessary

Screen: Yubikey registration (YubiKey Flow)
User action: press YubiKey button, a code appears on screen
Status: already existing, without adaptation necessary

Screen: SMS registration (SMS flow)
User action: fill in phone number and press button
Status: already existing, without adaptation necessary



Screen: receive SMS (SMS flow)
User action: fill in received SMS code on website
Status: already existing, without adaptation necessary