Vrije Universiteit Amsterdam

Master Thesis

Data portability on social networking sites

Author: Julien Benistant (2543043)

Supervisor: Sieuwert van Otterloo
Second reader: Davide Ceolin

A thesis submitted in fulfillment of the requirements for the VU Master of Science degree in Information Sciences

June 21, 2023

To my late mother. Her wish was for her children to achieve a degree in higher education. I wish you could have been here to witness this.

To my wife. Thank you for your continuous support even during the busiest and most difficult of times. This would not have been possible without you.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Sieuwert van Otterloo, for his invaluable guidance and support throughout the master thesis research. His willingness to help a struggling student for an extended period of time has made it possible for me to do the required research for and writing of the master thesis.

Abstract. With the General Data Protection Regulation (GDPR) in effect, social networking sites (SNSs) operating in the European Union must implement the right to data portabality. The right to data portability states that consumers should be able to export and import their personal data and be able to request direct transfer of their data to a different controller. The purpose of the thesis is to analyze data portability options offered by the most used SNSs, how data portability is facilitated, if the approach to data portability by sites is in line with the GDPR and if recommendations can be made in this regard.

1 Introduction

Social networking sites (SNSs) bring more and more attention to the issue of privacy and the protections of personal data due to their rapid growth rate and the daily use by many consumers [8]. As there are many users in different regions of the world, SNSs have to deal with laws and regulations which apply to these regions. Within the European Union, the General Data Protection Regulation (GDPR) is the leading regulation when it comes to the rights of consumers concerning their privacy and has been in effect since May 25 2018.

The purpose of the GDPR is to protect personal data and strengthen constitutionary rights of SNS users in a society which is becoming increasingly more digital [1].[1] Examples of constitutionary rights regarding data protection might be (but are not limited to) the right to have your data deleted from organizations or to have an organization move or send your data to a different organization before deleting it. The GDPR defines personal data as "any information relating to an identified or identifiable natural person" (Article 4, 1 of the GDPR) [1].

Article 20 of the GDPR entails the right to data portability, which will be researched on SNSs in the study. Data portability is the ability to to move data between applications or platforms [24]. It is oriented towards an interconnection of all digital services [10] and grants data subjects (such as site users) the right to transfer their personal data [18].

In this study, it is attempted to research if a common definition exists for an SNS, if and how data portability is facilitated by different websites, if the current approach to data portability in general is in line with article 20 of the GDPR, the right to data portability, or if a different approach might be more beneficial. This is done by examining different websites that are commonly used or are alternatives to such commonly used sites, and implementing existing difficulties in the mockup experiment. Also, an existing data portability platform is tested in the form of a technical experiment.

This van be formulated into the following research questions:

- Is there a common definition for a social networking site?
- How is data portability facilitated by different sites?
- Is the approach to data portability in line with the Right to Data Portability (RtDP)?
- Does the RtDP reach its intended purpose?

In the thesis, the following sections are presented. The initial approach is explained first, followed by the literature review. Then, the methods are described, and the results and analysis are presented. Next, the challenges are discussed, and, finally, the conclusions are drawn.

[1] `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02016R0679-20160504#tocId7`

2 Approach

2.1 Social networking sites

To determine which SNSs will be examined, a definition of a social networking site is required. Based on this definition, a selection of most used SNSs is made.

Definition of a social networking site

A social networking site is a site with the intent of connecting people to each other by means of a public profile, viewing connections, and social interaction [6] [19]. The aim of the SNS user is to interact with its social group in an online community [29]. While the definition by boyd & Ellison [6] and Kaplan & Haenlein [19] were the main purposes of using an SNS in 2007, SNSs have evolved to focus more on a dynamic form of content provided by the user, the groups that the user is part of, by connections (such as tags in photographs) and/or by the system (e.g. third-party sites) [13]. In this thesis, the latter purpose is considered an improvement by being more accurate. An example of a difference could be eBay or Marktplaats where social interaction is possible (older definition), but interaction with streams of content generated by connections is not possible (newer definition). The definition below given by Ellison & boyd [13] is adopted for this thesis:

A social networking site is a networked communication platform where:

- *users have unique profiles that consist of content provided by the user;*
- *users can publicly articulate connections that can be viewed and roamed by others;*
- *users can consume, produce, and/or interact with streams of content generated by users that are part of their connections on the SNS [13].*

The definition entails that users can articulate connections that can be viewed and roamed by others [13]. An example would be viewing the friends of a personal Facebook friend, or the user viewing the followers of a Twitter page that the user is following. While this fits in the sense of a network being a web of people/organizations being connected to each other through other people/organizations, a quick Google search on "what is a network" shows an alternative definition of a network: interact with others to exchange information and develop professional or social contacts. In this sense, any site which allows you to create a profile, view the profiles of others, and interact with others, could be considered a social networking site. Sites such as eBay and dating sites could then also be considered social networking sites. In terms of the right to data portability, such sites are considered relevant for the research of the thesis as the users have the right to control their provided data as described by article 20 of the GDPR [1]. Different considerations and the relatively old age of the articles from 2007 [6] and 2013 [13] make it difficult to construct a proper and definitive definition of an SNS.

Examined websites

The websites in table 1 below were examined to determine whether they are social networking sites using the three criteria of our definition of an SNS. Some SNSs were selected based on their focus of being an alternative and/or a competitor to another communication platform with similar functionality. The platforms with an asterisk (*) are alternatives for the first platform before without an asterisk. It should be noted that transferring data from LinkedIn to Twitter falls within the right to data portability, but it cannot be expected that the work experience of the user is transferred to Twitter, since Twitter does not accomodate this type of information.

SNS	Unique profile with content by user	User can have connections that can be viewed by others	Users consume/produce (/interact) with content generated by connections
Facebook	Yes	Yes	Yes
Sociall*	Yes	Yes	Yes
MeWe*	Yes	Yes	Yes
Instagram	Yes	Yes	Yes
Tookapic*	Yes	Yes	Yes
Pixelfed*	Yes	Yes	Yes
Twitter	Yes	Yes	Yes
Mastodon*	Yes	Yes	Yes
Movim*	Yes	Yes	Yes
Micro.blog*	Yes	Yes	Yes
TikTok	Yes	Yes	Yes
Dubsmash*	Yes	Yes	Yes
Likee*	Yes	Yes	Yes
LinkedIn	Yes	Yes	Yes
XING*	Yes	Yes	Yes
Opportunity*	Yes	No	No
eBay	Yes	No	No
Marktplaats*	Yes	No	No
Tinder	Yes	No	Yes
Parship*	Yes	No	Yes

Table 1. Examined websites including GDPR. art. 20 criteria

The alternatives were selected based on different online sources, such as ethical.net[2], makeawebsite-hub.com[3], careeraddict.com[4], technicalustad.com[5], makeuseof.com[6], and tomsguide.com[7].

Facebook direct transfer target sites

The Facebook target sites for direct transfer were also examined to determine whether they are SNSs according to the three criteria of our definition of an SNS and is shown in table 2.

For all target sites, one of the three criteria is met. According to the definition described in 2.1, this means that all of the target sites are not considered SNSs. It is interesting to note that Facebook has put in considerable effort to make data transfer possible to ten different target sites, but none of the effort is directed towards different SNSs.

2.2 Right to data portability

The right to data portability gives consumers the right to ask an organization for their personal data, which have been supplied by the consumer at an earlier time, and/or the right to ask an organization to transfer their data to a different organization [1]. In other words, it facilitates the reuse of personal data among data controllers, such as SNSs, by establishing a general-purpose control mechanism to be applied

[2] `https://ethical.net/guide/facebook-alternatives-guide-how-and-why-to-avoid-facebook/`
[3] `https://makeawebsitehub.com/facebook-alternatives/`
[4] `https://www.careeraddict.com/6-alternative-websites-to-linkedin`
[5] `https://technicalustad.com/linkedin-alternatives/`
[6] `https://www.makeuseof.com/best-tiktok-alternatives/`
[7] `https://www.tomsguide.com/best-picks/best-dating-apps`

SNS	Unique profile with content by user	User can have connections that can be viewed by others	Users consume/produce (/interact) with content generated by connections
Google Photos	Yes	No	No
Google Docs	Yes	No	No
Dropbox	Yes	No	No
Koofr	Yes	No	No
Google Calendar	Yes	No	No
Daybook	Yes	No	No
Photobucket	Yes	No	No
Blogger	Yes	No	No
Wordpress.com/Jetpack	Yes	No	No
Backblaze B2	Yes	No	No

Table 2. Examined Facebook direct transfer target sites including GDPR. art. 20 criteria

horizontally among sites [17], and grants more control over their personal data [5] [7].

Article 20, the right to data portability, appears as follows in the GDPR:

1. *The data subject shall have the right to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format and have the right to transmit those data to another controller without hindrance from the controller to which the personal data have been provided, where:*

   (a) *the processing is based on consent pursuant to point (a) of Article 6(1) or point (a) of Article 9(2) or on a contract pursuant to point (b) of Article 6(1); and*

   (b) *the processing is carried out by automated means.*

2. *In exercising his or her right to data portability pursuant to paragraph 1, the data subject shall have the right to have the personal data transmitted directly from one controller to another, where technically feasible.*

3. *The exercise of the right referred to in paragraph 1 of this Article shall be without prejudice to Article 17. That right shall not apply to processing necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller.*

4. *The right referred to in paragraph 1 shall not adversely affect the rights and freedoms of others.*

Paragraph 1 states that the data subject, i.e. the website user, has the right to receive the personal data that was provided to the data controller, i.e. the SNS, in a structured, commonly-used and machine-readable format. Since it is a right, the SNS user has to request the personal data from the SNS. If the SNS user receives the personal data, it may be used by the user on other SNSs. This method of data transfer through the user will be referred to as *indirect transfer* in this thesis.

Paragraph 2 states that if the user acts on the right of data portability in accordance with the right described in paragraph 1, the user has the right to have the personal data directly transmitted from one controller to another. This right is applicable to SNSs by merely switching out the word "controller" with SNS. This method of data transfer will be referred to as *direct transfer*. The purpose of this right is to

counteract switching costs for SNS users (data subjects) and to increase competition between SNSs (data controllers) [20]. It is also stated that the right to direct transfer is applicable where technically feasible. Even before the GDPR was finalized, there were doubts regarding the implementation [9] as compliance costs may be too large for smaller organizations [3] and is a considerable obstacle [24]. The implementation might require a high volume of work and money depending on the circumstance and the amount of code to be implemented [2]. Nonetheless, the right applies for both a start-up software company and a monopolist [14]. Therefore, enforcing this right may incur disproportionate costs and efforts [11].

Paragraphs 3 and 4 state that the abovementioned rights or excercising those rights shall be without prejudice, meaning without change or harm, to article 17 (the right to erasure) [1], and that it should not affect other data subjects (SNS users). These paragraphs do not state any further rights regarding the right to data portability.

The right to data portability and the GDPR in general do not describe or define the term *social networking site*. The GDPR does describe that the right to data portability is applied when the processing of the data of a person is done with permission or by agreement and is performed by automated means [1]. If these criteria are not met, the data subject could request a data transfer, but the data controller would not be obligated to honour the request.

Recital 68: Right to Data Portability

Recital 68 of the GDPR[8] concerns article 20 [1] and allows organizations to learn when and how to comply with it.[9]

What is further clarified in the recital is that the right to data portability should not apply to where data processing is based on legal ground other than consent or contract, and that the right should not be exercised against controllers that process data because of public duties or where processing of personal is necessary for compliance with a legal obligation [1].

The data subject right to reveice their personal data should not create an obligation to for data controllers to implement or maintain technically compatible processing systems [1].

If a data subject receives personal data that also concerns another data subject, it should be without change or harm to the other data subject [1].

Machine-readable format

Article 20 of the GDPR states that the data download should be provided in a structured, commonly used and machine-readable format [1], but does not specify which formats are considered machine-readable or how it should be interpreted [27]. Therefore, the scope of "machine-readable" will be dependent on the interpretation of the implementer [28]. Filetypes that are commonly used in a specific domain which allow for export and import, also fit the criteria. An example could be financial systems using a standard for data export and import. And while HTML is not always considered a machine-readable format in different studies, many sites use the format for personal data exports. With this in mind, it cannot be stated for any SNS would not adhere to the GDPR requirement of offering the data in a machine-readable format, as there is no definitive specification as to what is considered a machine-readable format. It is, however, necessary for the thesis to have a clear point of view with regard to whether HTML is considered

[8] `https://www.privacy-regulation.eu/en/recital-68-GDPR.htm`

[9] `https://www.americanbar.org/groups/litigation/committees/minority-trial-lawyer/practice/2019/a-very-brief-introduction-to-the-gdpr-recitals/`

a machine-readable format.

Where previous studies [4] [26] have classified HTML as a form of a machine-readable format, next to XML, CSV, and JSON, or have classified it as ambiguous [30], this study does not classify HTML as such. HTML is a markup language which has its main focus on how to display a page to a human user and not how the information itself is displayed. And while HTML might be considered useful as a machine-readable format when some form of tagging or metadata is used, the SNSs which provided HTML downloads did not provide such tags or metadata.[10]

2.3 Data portability by third parties

The right to data portability appears to focus on the users requesting their personal data to be transferred by the data controller, in this case SNSs. However, there are third party applications that are able to store your personal data to be used at a later time on other sites. In such a scenario, the third party would be able to provide the personal data for a newly created account on a new SNS. Such third parties are also referred to as Personal Data Spaces (PDSs) [22].

An example of a PDS providing a solution for data portability, is Meeco.[11] Meeco attempts to enable a personal data ecosystem that includes people participating directly in the value chain via the 'API-of-Me'.[12]

3 Literature review

3.1 Case studies

Since the GDPR came into effect, the implementation of data portability has been examined in multiple studies, focusing on different types of organizations.

Sørum & Presthus examine the right to data portability in practice by testing the extent of data portability offered by different physical shops, online shops, social networks and a search engine. [26] The study relates very well with the research in this thesis, as an attempt is made to test data portability on social networking sites.

A case study on data portability exports was done by Barth in IoT platforms in 2021 [4]. In the study, data portability exports were examined as it is an expected service based on the first right to data portability described by the GDPR [1].

The study by Syrmoudis et al. was published in March 2021 and examines the data export and data import capabilities of 190 online services and whether specific industry sectors are more effective in enabling data portability [27]. Syrmoudis et al. name examples such as social networks, map or fitness applications as key online services. Of the 182 online services that were sent data export requests, 52 fulfilled the requirements of the GDPR including the timeframe in which the data export should be sent [27]. Also, none of the 190 online services directly offer to import data generated by a right to data portability request (i.e. data export), while 11 offer minimal import possibilities (e.g. survey service imports contacts but no surveys), 20 offer partial import (import for some but not all core functions), and 13 offer full data import (import possibilities for all core functions) [27]. It should be noted that Syrmoudis et al. do not offer a formal definition of an online service. Social network sites are named as a key online service [27],

[10] `https://www.data.gov/developers/blog/primer-machine-readability-online-documents-and-data`
[11] `https://www.meeco.me/`
[12] `https://docs.meeco.me/`

and while SNSs could be categorized as online services in a general sense, this study has a more specific definition where SNS users have unique profiles with user-generated content and can consume, produce and/or interact with content generatured by other users.

A study by Kuebler-Wachendorff et al. from 2021 finds that the majority of services in their right to data portability research do not provide any data import options [21].
These case studies examine data portability based on the export and import of data and not by means of direct transfer. It would appear that studies involving direct data transfer and its implementation are not widely tested yet based on the performed literature study. With regard to indirect data transfer, it might be interesting to examine if there are significant differences between the results this study in 2021 and the results of the case studies.

3.2 Facebook whitepaper

Facebook has a whitepaper concercing data portability, which attempts to outline the data portability challenge, to determine what data portability is, which and whose data should be portable, and privacy concerns which may rise a result of data being transferred [12].

Challenge

In the challenge, it is stated that Facebook has offered "Download Your Information" since 2010, which was later improved to adhere to the GDPR and the California Consumer Privacy Act by, for example, having the data delivered in JSON format.[13] The challenge itself refers to direct data transfer and the collaboration on the Data Transfer Project (DTP) with Google, Microsoft, Twitter, Apple and others [12]. More details on the DTP and its current state are mentioned in section 3.3 of the thesis.

Data portability

When trying to define data portability, the whitepaper mentions three types of user-directed data transfers: open transfers, conditioned transfers, and partnership transfers. Open transfer would be a scenario where there are no restrictions from the source site on the target site [12]. Conditioned transfer would be a transfer method in which the source site has specific conditions a target site should meet in order to receive the data where the sole purpose is receiving personal data [12]. Partnership transfer is described as a transfer where the source site and the target site have a larger relationship beyond personal data, such as integrating features from one site on a product of the other site [12]. The whitepaper contemplates whether conditioned transfer would fall under the GDPR [12]. As the right to data portability mentions "the personal data" without further specification, the uncertainty in the paper is somewhat understandable. However, as there is no further specification, would it not mean that it entails all personal data?

Tracked data

The whitepaper discusses that it is less clear what data, other than user-provided data, should be included [12]. As paragraph 1 of article 20 the GDPR states that it concerns data which "he or she has provided to a controller" [1], the rest of this discussed question which focuses on observed and/or inferred data is not relevant in the context of the GDPR, which is researched in this thesis.

Data ownership

[13] `https://oag.ca.gov/privacy/ccpa`

Facebook discusses data ownership with examples such as a person uploading a video of herself and some friends [12]. The question is posed whether the other people in the video should have rights regarding the video. While this research does not go in-depth to this extent, the issue of images from shared posts being transferred does come up in the tests performed in this study.

Privacy vs. portability

It is also noted that there is little guidance in protecting privacy while laws come into effect regarding data portability [12]. This study has a more direct focus on the adherence to the GDPR and the likeliness for data transfer to be used, without zooming in on privacy concerns. The regard for privacy should be considered important for companies handling large amounts of personal data.

Responsibility after transfer

The Facebook whitepaper asks about who is responsible for the data after transfer [12]. While it can be considered an interesting topic to study, the thesis does not dive into the responsibility of the data after transfer.

3.3 Data Transfer Project whitepaper

The Data Transfer Project is a collaborative data transfer project initiated by Google, Microsoft, Facebook, and Twitter. Since its inception, other tech corporations, such as Apple, have joined the DTP. On July 20, 2018, a white paper was released by the DTP team, which provides a more in-depth understanding of the project and its details [15]. It is a serious tool, which is updated over time and is used by the involved parties. So, for example, all direct data transfer possibilities that are offered by Facebook work with the DTP.

The whitepaper is titled *Data Transfer Project Overview and Fundamentals* and addresses the principles of the DTP, its focus on direct data transfer from one provider to another, architectural constraints, system components, deployment of a DTP instance, security and privacy of user data, ecosystem issues that the project faces, implementation of the project, and ways to participate [15].

The goal of the DTP is to establish an open-source ecosystem for the development of tools and techniques to facilitate portability of specific user data between providers (data controllers) [15].

4 Methods

To reach the set out research goals mentioned in section 1, multiple research methods are used. Different websites, including SNSs, were examined and tested. Exporting data and attempting to import data was done with the existing accounts of the author and sites for which the author did not have an account yet, new accounts were created. Both full and partial data export and import were examined since not all sites have the same type of export available. Direct data transfer was also tested by selecting the full profile and by selecting a subset of the data to be transferred. Next, the mockup experiment with an accompanying survey were held.

4.1 Selecting sites to study

A list of most used social networking sites is used to determine which SNSs are examined. As most lists cover social media use, which entails all types of media and not just SNSs, such a list will be examined to determine which SNSs, based on the definition of an SNS, are most popular. These are Facebook,

LinkedIn, Instagram, TikTok and Twitter.[14]

Next to most used social networking sites, different networking sites for specific purposes were considered to be examined if they would fit with the definition described above. Examples such as dating sites and consumer to consumer selling sites raise the question to what extent a social networking site should be defined in terms of degrees of connections. Regarding the GDPR, each of these sites should adhere to the right to data portability as they handle user-provided personal data, and is considered a relevant site for the study.

The selected sites will be tested by examining the site for data download options and actually downloading the data to discover if it works as expected. The sites that offer data upload options will also have the data upload options tested to examine if it works as expected. A study from 2016 states that transferring a profile to a competing service requires time and effort, since the data is not compatible enough between sites [14]. It will be interesting to examine if there have been improvements in this regard over time.

4.2 Mockup experiment and survey

The mockup experiment is based on site reviews, where difficulty was noticed in discovering data portability options. The experiment leads the participant through the process of three different methods of data portability: indirect transfer, direct transfer where the source site is sending the data to a target site, and direct transfer where the target site retrieves data from a source site. This last method will be referred to as the *OAuth* method in the thesis, since it requires the user through the target site to sign in to the source site. In short, we name them: indirect transfer (download/upload), direct transfer (data push), and OAuth transfer (data pull).

Afterwards, the participant is asked to fill in a survey which enquires about preferences regarding the data portability methods, its ease of understanding, and its ease of use. Some parts of the mockup experiment contain errors in the result of the data transfer. The errors were built into the experiment based on erroneous results from personal experience. For instance, when selecting to transfer a subset of images from Facebook to Google Photos, all images of the author were transferred. This type of error was used in the mockup experiment to transfer all posts when the option was selected for posts to not be sent (Appendix B option figure 30 and resulting figure 33). The survey contains questions about the extent to which the participants noticed the errors and how they would feel about the errors if they were to experience them on a real site. As you can see in appendix C, the survey contains four questions relating to data transfer errors. If users are unaware of data transfer errors, wrong or incomplete data might get transferred, which in turn might result in loss of data when, for instance, the profile on the source site is deleted.

In the survey, different profiles were used for the different sections to deal with different situations, such as special characters in names. It should be noted, however, that it could have been better for research purposes to keep de profile screens the same, as it would be a comparable testscenario starting position to the other parts of the experiment.

Besides questions about the mockup experiment, the survey aims to determine the awareness of the GDPR and the awareness of the right to data portability of the participant. This is done by asking the participants if they have previously heard of the GDPR and the right to data portability and how they became aware.

When looking for data portability options on Facebook, it is unfortunate that it takes a relatively long time to discover them. The hope was to easily find the options and to test them out. And while Facebook might consider it a function less used than their other site functions, it is interesting to examine how users

[14] `https://onemedia.nl/social-media/`

respond to the process of searching these options. For the mockup experiment, an attempt was made to simulate the relatively large amounts of clicks required to reach the desired functionality. It might be interesting to examine, as other research suggests that mostly technologically interested and compentent site users have the desire to swith between services using data portability [23].

Initially, there were 21 participants who performed the experiment and subsequently completed the survey. Parts of the results are based on these initial participants. At a later moment, the experiment was adjusted and an additional 10 participants completed the experiment and the survey. These 10 responses will make up a different part of the responses. In parts where the experiment was not adjusted, the total amount of 33 responses will be examined.

In one case, the respondent was unable to complete the experiment due to an issue where the next page/step would not load. While no other respondents had mentioned this error, by not completing the experiment, there is no way to start the survey. Since the site had no form of tracking built in, it was not possible to measure if respondents had stopped participation during the experiment.

4.3 Technical experiment

It can be assumed that smaller organization will want to keep the cost for data portability options low. One way of reducing work is to connect to an existing platform which facilitates data portability.

How can organizations connect with such an existing data portability platform? With this technical experiment, the demo of the Data Transfer Project is tested with Docker locally, and the structure of the open-source project is examined. The purpose of this experiment is to attain a sense of the difficulty of connecting with an existing data portability platform. The difficulty will be assessed based on the four year development experience of the author. It should be noted that the sites in the demo are working on the live sites.

Setup

For the technical experiment, the DTP was run locally.[15] Due to time constraints, Docker was used to run the demo image of the DTP.[16] The alternative to Docker would be to download the code and follow the steps in the documentation to get it working, which would require more time to set up. Both approaches lead to the same working demo application. The difference is that the code can be edited and/or expanded for further testing purposes, which is not possible with Docker.

The documentation states that the DTP can be run locally via Docker or via code. Docker was chosen for the experiment, as there exists a working Docker image, and it was expected to cost less time than downloading the code and going through the steps to get the different components working. The documented steps to run the DTP locally[17] were followed.[18] The app was started relatively quickly and by browsing to the frontend of the app, the demo appeared.

Expectations

Users may have expectations with regard to the resulting outcome of the direct data transfer. Based on the observations of the experiment, it might be possible to determine if user expectations are met.

[15] `https://github.com/google/data-transfer-project/blob/master/Documentation/RunningLocally.md`
[16] `https://www.docker.com/`
[17] `https://github.com/google/data-transfer-project/blob/master/Documentation/RunningLocally.md`
[18] `https://github.com/google/data-transfer-project/blob/master/Documentation/Keys.md`

5 Results and analysis

5.1 Social networking sites

Table 3 shows a table of the examined communication platforms' data download options. Table 4 shows a table of data upload options. Both tables show the download formats and remarks regarding data download or upload. In addition to data download and data upload, this section will also examine direct data transfer options on Facebook.

Data download

As seen in section 2.2 of the thesis, paragraph 1 of article 20 of the GDPR states: *The data subject shall have the right to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format.* This statement is made more clear with the information gathered in the thesis regarding machine-readable formats. It is, however, not specified further outside of the thesis, what formats fall under machine-readable formats.

Table 3 shows that some SNSs do not provide the option to download one or more files containing the personal data of the user, which is illustrated by showing a hyphen (-). Where automated download options were not found, the sites were contacted through email or a messaging system to request a personal data export. For the sites where an answer is still pending, the word **Pending** is placed instead of a hyphen or download format. Of the fifteen SNSs that have made data export available, ten offered the download in one or multiple files in a machine-readable format which have been highlighted in the figure with a green font colour. Other download formats, that are not considered machine-readable formats according to this thesis, have a red font colour. This means that eleven of the 21 examined SNSs do not have an automated process of downloading personal data available.

Figure 1 shows the popularity of the different data download formats based on table 3. JSON is the most prevalent, with HTML being a close second.

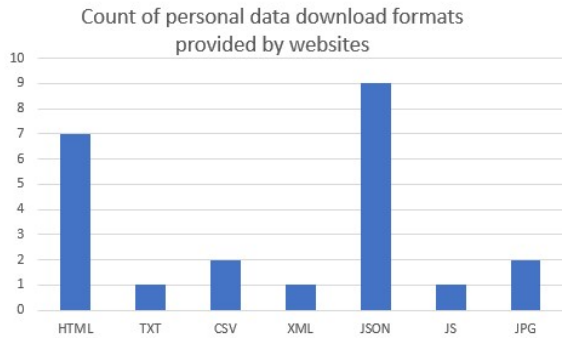

Fig. 1. Count of the provided data download formats based on table 3

Data upload

Compared to data file downloads, a smaller number of SNSs, two out of 21, offer the functionality of uploading your personal data through file upload on the site as is shown in table 4. The SNSs that do not offer the functionality are illustrated with a hyphen (-) in table 4. It should be noted that with both

Social networking site	Download format	Remarks
Facebook	HTML, **JSON**	Images are also downloaded in JPG format.
Sociall	HTML, **JSON**	
MeWe	TXT	
Instagram	HTML, **JSON**	Images are also downloaded in JPG format.
Tookapic	JPG	Other than posted images, personal data cannot be downloaded.
Pixelfed	**JSON**	
Twitter	**JS**	JS files contain single variable with JSON body.
Mastodon	**CSV**, **JSON**	
Movim	**Pending**	
Micro.blog	BAR, **JSON**, **XML**	BAR and XML (WXR) downloads are for posts only. JSON download is for profile information and the replies the user has made.
TikTok	TXT, **JSON**	Can only be requested with the app.
Triller	**JSON**, JPG	
Dubsmash	-	No contact options for the Chinese owner of Dubsmash were found.
Likee	**Pending**	
LinkedIn	**CSV**	
XING	HTML	
Opportunity	**Pending**	
eBay	HTML	
Marktplaats	HTML	By visiting https://privacy.adevinta.com/marktplaats.nl
Tinder	HTML, **JSON**, HTML	By visiting https://account.gotinder.com/
Parship	**Pending**	

Table 3. Social networking sites personal data download option

SNSs not all types of personal data can be uploaded with a machine-readable file.

Social networking site	Upload format	Remarks
Facebook	-	
Sociall	-	
MeWe	-	
Instagram	-	
Tookapic	-	
Pixelfed	-	
Twitter	-	
Mastodon	CSV	Upload only possible for specific features.
Movim	-	
Micro.blog	XML, ZIP, JSON	XML (WXR) uploads are meant for WordPress posts, ZIP uploads are for Medium and Tumblr, and JSON is for Ghost blogs.
TikTok	-	
Triller	-	
Dubsmash	-	
Likee	-	
LinkedIn	-	
XING	-	
Opportunity	-	
eBay	-	
Marktplaats	-	
Tinder	-	
Parship	-	

Table 4. Social networking sites personal data upload option

Data upload test

Since data upload is offered on Mastodon and Micro.blog, both Twitter alternatives, a reupload of downloaded personal data was performed to test the functionality of the automated data upload feature. Both sites allow post creation, image upload, video upload, reply to/comment on posts, and bookmark posts. Differences with Twitter are that both sites do not allow the user to "like" posts, and that they allow bookmarks where Twitter does not.

Mastodon

While Mastodon allows specific lists of information to be downloaded in CSV format, and posts and uploaded media in JSON and JPEG respectively (Appendix A figure 9) by using the "request archive" functionality, they only allow for the upload of specific information in CSV format (Appendix A figure 10). The upload of the JSON files provided in the archive is not possible at the moment of writing this thesis.

After downloading a list of followed users (named *Follows*) containing three users, a newly downloaded list was reduced from three to two by unfollowing one user before the download. When the CSV file was uploaded containing the list of three users, in both the merge and overwrite scenarios the list increased to three users. When uploading a smaller list to overwrite, the functionality worked as expected: Only the uploaded smaller list of users was present and active. The upload works as expected, as the correct (links

to) users and correct numbers of users were shown. With all CSV upload options, the uploads worked as expected. And while it may not be as extensive as a full personal data collection, it is a first step in making personal data uploads possible.

It is interesting to notice that Mastodon has split the download/upload option into several categories, as it may provide more control in different types of information for the users, the data subjects, in which data they wish to transer.

All CSV upload options worked as expected with the downloaded personal data from Mastodon in CSV format.

Micro.blog

According to the file uploads and downloads of Micro.blog, the file formats that match are XML(WXR=WordPress format) and JSON. Micro.blog allows for a download of your posts in WXR format according to the website (Appendix A figure 11). In the upload section of the website, both WXR and regular XML can be used for the upload of posts. After the upload was completed, the exact same posts appeared with the same text and with the same creation date. This functionality worked as expected. (Appendix A figure 12)

Table 5 shows a quick overview of the Micro.blog results. The upload section of Micro.blog allows for the upload of JSON formats from Ghost[19] (Appendix A figure 13) and from Foursquare/Swarm[20] (Appendix A figure 14). While a prompt appeared saying that the file was uploaded, uploading the downloaded JSON file from Micro.blog had no effect on the account or the posts. In this case, the upload option was offered, but the result was not as expected. The unsuccessful upload of the JSON file might be because of the mismatch in serialization between the uploaded JSON file and the JSON serialization that was expected by the site.

The unsuccessful upload of the JSON file is expected to be due to it being a serialization file type. This means that there are more

Upload type	Worked as expected	Remarks
XML posts from Micro.blog	Yes	The exact same posts with the same creation dates appeared on the profile.
JSON from Micro.blog	No	Nothing changed on the site.
JSON from Ghost	No	Nothing changed on the site.
JSON from FourSquare\Swarm	No	Nothing changed on the site.
Tookapic*	No	Nothing changed on the site.

Table 5. Upload types to Micro.blog based on file type and source site

It is interesting to notice that the upload of posts is possible in XML format, which works as expected when uploading a file downloaded from Micro.blog itself. It is a different challenge, however, to get file uploads in other file types and from other sites working properly, as the offered functionality did not offer the expected results.

[19] `https://ghost.org/`
[20] `https://www.swarmapp.com/`

Direct transfer test on Facebook

This section examines the direct data transfer options and examines the results of the test transfers initiated on Facebook. Facebook offers direct data transfer to 10 different target sites, which are shown in figure 2. Five types of file transfers are specified: photos, videos, notes, posts, and events (figure 2). As photos and videos are comprised of the same list of target sites and the transfer page to these sites show options for both, they have been grouped together as the *photos/videos* type. Also, Facebook offers transfer type *notes*, which is a type of post that can no longer be created. Since the used source account has no notes, they have been left out of this research. Since all *notes* targets are also *posts* targets, the total list of target sites remains the same.

It should be noted in figure 2, that different (one or more) filetypes can be sent to different target sites. While viewing the list, it can be observed that the sites are not (fully) comparable or similar to source site Facebook.

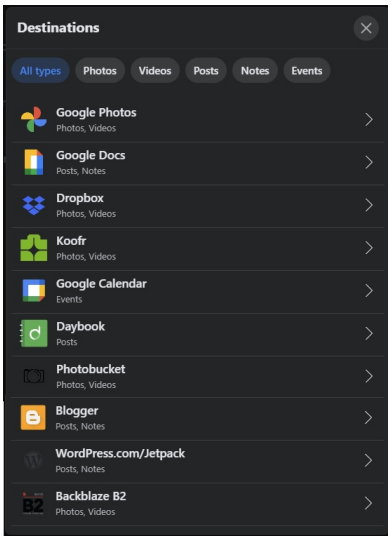

Fig. 2. All data transfer destinations offered by Facebook

Facebook direct transfer success

Table 6 gives an overview of the different transfer types Facebook direct transfer offers and whether the selected range of data resulted in a successful or unsuccessful transfer. The data transfer results were equal for all target sites, which allows for the assumption to me made that it is related to the data collection of the sending party (Facebook). In the case of Photos/Videos, the date range transfer is unsuccessful, because photos/videos outside of the date range were also transferred. With Events, no events were transferred to the target site(s).

With all types of transfer, selecting the *none* option, resulted in no data being transferred, which can be translated as the selected data transfer option to be successful. It could be discussed whether such an option should be available when a user is attempting to transfer data.

Transfer type	Transfer all data	Transfer date range	Transfer specific data	None
Photos/Videos	Successful	Unsuccessful	Successful	Successful
Posts	Successful	Successful	N/A	Successful
Events	Unsuccessful	Unsuccessful	N/A	Successful

Table 6. Facebook direct transfer results per transfer type

Below are the more detailed findings of the performed tests per file transfer type.

Photos/Videos

Figure 3 shows the list of target sites Facebook has for photos/videos transfers.

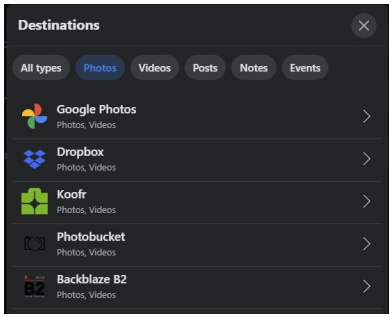

Fig. 3. Photo/Video data transfer destinations offered by Facebook (January 2023)

While testing the different options for the transfers, the same sets of photos and videos were sent to the target sites by Facebook. The assumption can be made that Facebook first selects the files based on the specified user criteria and subsequently enters a process sending the files to the target site. All target sites for photos and videos received the same data, data selection seems more of a factor for correct data transfer.

For data selection, Facebook offers the following options, which are displayed in two similar lists on the same screen (alternative shown in brackets) :

- All your photos [videos]
- Select date range
- Specific albums... [videos...]
- None

When choosing *All your photos* or *All your videos*, all files appeared to get transferred.

When a specific photo album or a specific video was selected, these were also transferred correctly.

The *none* option did not send any files, as was expected.

The *date range* option showed unexpected behaviour. For the transfers, different starting dates in October 2019 were selected and end date October 1st, 2022. In all cases, images from as early as 2014 were

transferred, which do not fall within the specified range.

Posts

Figure 4 shows the list of target sites Facebook has for posts transfers.

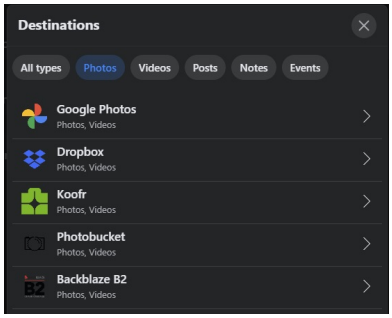

Fig. 4. Posts data transfer destinations offered by Facebook

For data selection, Facebook offers the following options for posts:

- All your posts
- Select date range
- None

During testing, date range appeared to work properly.

To Google Docs, the textual content was copied into a Google Doc and a different location was used for accompanying images.

On Daybook, the posts were placed on the specific diary pages accompanied by the images of the post.

When checking Blogger after transfer, the posts were displayed with the accompanying photos.

The transfer to Wordpress.com/Jetpack resulted in the Facebook posts becoming concept WordPress posts with the accompanying photos.

It is interesting to note that one of the transferred posts was a share of a post belonging to a different page. With this post, Facebook copied only the text of the post of the user making the request and not of the page that made the original post. It did, however, send the photos that were posted by the page. This was noticed on all resulting SNSs.

Events

Facebook only has Google Calendar as a target site for events transfers 2.
noindent For event selection, Facebook offers the following options for posts:

– All events
– Select date range
– None

After the completion notice on Facebook, no events are to be found when scrolling back in Google Calendar. It also worth noting that, while calendars can be used to plan ahead, it is only possible to transfer data up until the current date.

5.2 Mockup experiment and survey

The mockup experiment is set up as a custom created series of web pages with an explanation bar on the right side of the screen. The first page the user lands on, is a logged in user page of a fictional person on a fictional SNS, named Connexus. The explanation on the side asks the user to try to download their personal information on the fictional site (Appendix B figure 15). The goal of part 1 of the experiment is to find the page where the data download option resides, download the data, view the data, login to a different fictional SNS named NEXTOP, and upload the data to the new site. Part 2 asks the user to transfer their data directly from Connexus to NEXTOP, and part 3 asks the user to create an account on NEXTOP and to retrieve the personal data from Connexus using OAuth.

Data transfer method preferences

Considering the preference in data transfer methods, in the survey, there were three specific options to choose from or a fourth option indicating that the participant has no preference. Figure 5 shows how the participants are divided over the options.

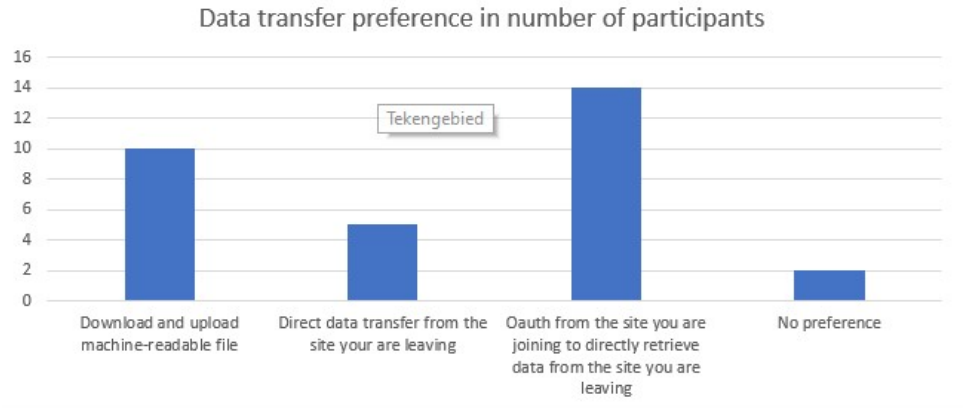

Fig. 5. Data transfer preference according to the survey

The options of downloading and uploading a file and using OAuth login from the site the user is joining, are the preferred options of the survey participants with 32.26% and 45.16% respectively. The participants that preferred the file download and upload, generally accounted it to having more control over the data and safety concerns. The participants that preferred the OAuth option, mostly attributed it to being the easiest, most convenient option.

When the participants were asked if it was clear to them which transfer options were available on real websites, 14 out of 31 disagreed or strongly disagreed, 6 neither agreed nor disagreed, 3 agreed, and

none strongly agreed. With the difficulty level of the mockup experiment, the general direction appears to skew towards not being able to understand where the options might reside on real sites. Since the mockup experiment is a relatively simple version of an SNS, the complexity of a real site might be considered more difficult to examine and traverse.

Impact of data transfer errors

As stated in the methods section, it is possible for errors to occur when transmitting data. Figure 6 shows the number of participants that initially noticed the errors built into the direct data transfer part of the experiment. The profile page, which is the starting point, can be viewed in Appendix B figure 27. On the transfer page (Appendix B figure 30) two options are shown to transfer:

- *Only profile information,* ***no posts***
- *All profile information,* ***including posts***

Only the first option is selectable, forcing the user to select a transfer without posts. The resulting page (Appendix B figure 33) shows that, unfortunately, both profile information and posts have been transferred.

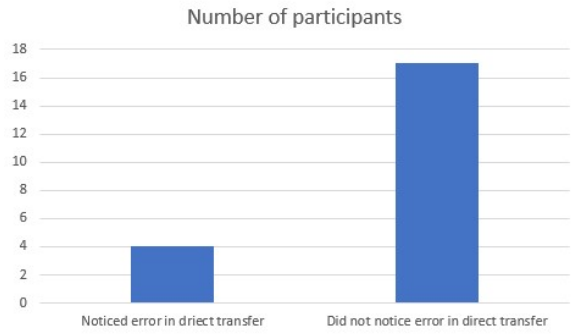

Fig. 6. Participants noticing transfer errors according to the initial survey

Most participants did not notice anything wrong right after the data transfer. Of the four participants that did notice errors, three named the error built into the direct data transfer part, the other participant named a difference resulting from part 1 (file download/upload) of the experiment. For the first part of the experiment, this participant would be partially correct, since none of the data was successfully transferred to the target site.

As most participants did not notice the errors, the comments were examined. Of the participants that did not notice any errors, most commented to be unaware that checking for errors was part of the experiment. Following up on this, the experiment was adjusted to make the participant aware during the experiment to check for errors and was sent out again. In the screens the added explanation is underlined. (Appendix B figures 27 to 33). The adjusted experiment had 10 respondents. Below is the resulting table of the adjusted experiment.

The results from the adjusted experiment in figure 7 show a greater percentage of respondents that noticed the errors (40%) than the respondents of the initial unadjusted experiment (19.05%) as shown in figure 6. While 10 respondents is not a large number of respondents, in percentage terms this group has more than double the amount of respondents that noticed the errors.

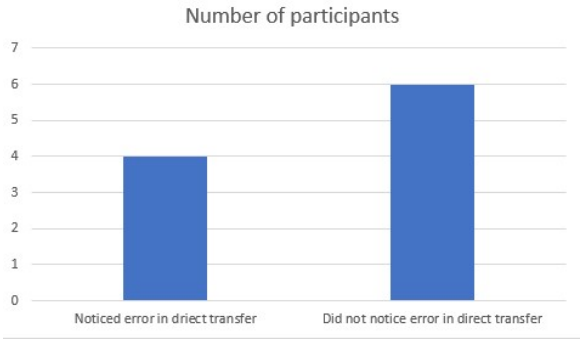

Fig. 7. Participants noticing transfer errors according to the adjusted survey

After learning about the errors that occurred, part of the 31 respondents changed their consideration to use direct data transfer. Figure 8 shows the likeliness of participants to use direct data transfer before and after learning about the transfer errors. It should be noted that the resulting 2 indecisive participants started as participants that were likely to use this method. Also, there are no cases where a participant started out as not likely and after learning about the errors became likely to use it.

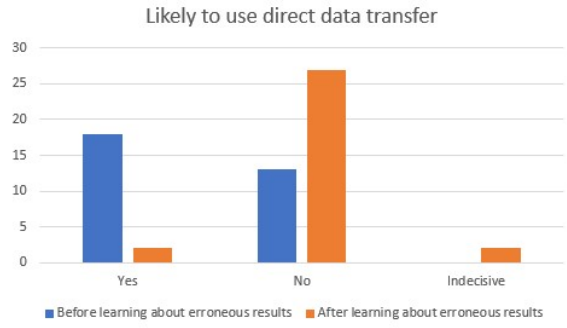

Fig. 8. Participants likely to use direct data transfer before and after learning about erroneous results

5.3 Technical experiment

Preparatory step

The DTP documentation describes the requirement of API keys to connect with the different sites, and provides links to the documentation of the sites on how to create the required IDs and secrets.[21] Without API Keys, the sites are not able to verify the user and are not able to establish a safe connection.

First, it was attempted to test the app without API keys. The frontend guides the user through the steps of choosing the datatype to be sent, and choosing the sending and receiving sites. Without an api set for the sending party, an error prompt appeared from the demo app, stating "Http failure response for https://localhost:3000/api/transfer: 500 OK".

[21] `https://github.com/google/data-transfer-project/blob/master/Documentation/Keys.md`

Next, the API keys for Facebook were created and set. The demo app was tested with different receivers without API keys and with Facebook as the sender. The result was that no error message prompt was shown by the demo app. Instead, the receiving sites showed an error message notifying the user in some form that the API keys were missing.

The following step was adding API keys to the senders. Most sites required a redirect URL for API keys to be generated. This was an issue as a local app cannot provide a valid redirect URL for the sites to refer to. The two exceptions were Facebook and Google Photos. Facebook did not request a redirect URL, and depending on the type of the application Google Photos might ask for a redirect URL. By choosing the option "desktop application" for the API keys to work with, Google Photos did not request a redirect URL and provided the required keys.

Google Photos is a photo sharing and storage service and is not considered an SNS. For the technical experiment, however, it was the only target site able to accept data (in the form of images) through a locally running DTP demp app.

Testing

With the API keys set for Facebook and Google Photos, it was attempted to send photos from Facebook to Google Photos. The demo app redirected to Facebook to log in, then redirected to Google to log in, and after some seconds the photos from Facebook were available in Google Photos. In comparison to the available direct transfer offered by Facebook towards Google Photos (see option in figure 3), the photos were not put in folders and were not the complete collection of photos from Facebook, but a transfer between two sites through a third-party app had occurred. It can be stated that some photos were transferred successfully. However, folder structures were not maintained, an unexpected set of photos was sent, and all sent photos were transferred together into the main folder. As folder structure is maintained when sending photos from the Facebook site to Google Photos, this behaviour was also expected from the demo app.

Expectations/requirements

Users might have expectations when it comes to data transfer. Maintaining folder structure could be considered a requirement for the transfer of photos. Other expectations might be that when a specific date range is selected, only photos from that date range are transferred, which did not occur with the direct data transfer from Facebook. The following list could be considered as expectations or requirements users can have in transferring photos:

- Photos are not lost in the transfer
- Photos are placed in the expected location
- Photos are transferred of selected date range
- Specific photo album is transferred
- Specific photo album (folder) is transferred
- Folder structure is maintained

These examples are considerations to take into account for websites that want to implement this type of data transfer.

Future work

It is possible to send photos from Facebook to Google Photos from the Facebook website. The difference between this process and the demo app is that with the demo app the developer is able to discover the basic structure of the app and think of methods to add a site to the DTP.

To add another site to the DTP, the source code will have to be expanded in addition to the steps performed in the technical experiment. From the viewpoint of a developer, knowledge of and experience with JSON and Java is required, and the structure of the source code will have to be examined thoroughly to not miss any necessary components of the DTP that need expansion. This would need to be researched further in a future study to make a more accurate assessment about adding a site to the DTP or taking a different path to achieve direct data transfer as described the right to data portability.

6 Discussion

6.1 Live sites

By examining live social networking sites, the situation of the site may change when it is studied at different moments. As one situation was studied and described, no screenshots were taken. By the time the thesis was examined, the situation on a site had changed and more data portability options were available.

6.2 Limitations

Technical issues during mockup experiment

The mockup experiment has been built using the basic methods of web development: HTML, CSS and JavaScript. While it was expected that these methods would not cause any technical issues for the participants, there was one recorded case where a mobile browser was unable to load a specific HTML page which was the page that led to the next step. While one participant actively provided this feedback, it might have occurred with more participants which in turn decided to quit without giving feedback.

Mockup experiment sample size

The experiment takes a varying amount of time, depending on the respondent, since the respondent has to examine a mockup site without active assistance. Because of the searching done by the respondent, this takes a larger amount of time. The accompanying survey was comprised of approximately 25 questions, of which most were open-ended, which take longer to answer and lead to larger item non-response [25]. While the search in the experiment may be more true to actual sites, it may have deterred respondents from continuing with the experiment and where the survey has the tool to examine where a respondent has stopped answering the survey, this feature was not available for the mockup site. A larger sample size might have been gained if the mentioned site complexity and survey time were reduced.

Mockup experiment time constraint

A relatively large amount of time was spent building multiple versions of the mockup experiment due to complex JavaScript implementation in the initial mockup. Due to the time expense, the mockup was not optimized to work on smartphones, while it worked on desktops, laptops and tablets. Smartphone users were better off turning their phone sideways. In hindsight, this has limited the amount of possible respondents, since most internet users access the internet through their smartphones.

Technical experiment time constraint

For the technical experiment, the amount of work to get the code up and running, as compared to loading and testing the docker file, would take considerably more time to achieve based on the programming skills of the author. If there was more time available to get the code running, experimenting with the code for a new site to connect with other DTP parties might have been a possibility worth exploring.

6.3 Adjustments

Mockup experiment adjustments

While attempting to retrieve information through data analysis on the mockup experiment survey, it became clear that most participants were unaware to check for data transfer errors. By adding specificity to the experiment explanations, results were collected from a new set of participants and is reflected in the results by comparing it with the initial survey results regarding the transfer errors. The final version with the improved instructions can be examined in Appendix B.

7 Conclusion

In this paper, article 20 of the GDPR, the right to data portability, was examined on different websites, including social networking sites. The aim was to learn the extent to which SNSs have implemented features into their platforms to facilitate the right to data portability and in what ways, if users are able to find and use data portability options on a mockup SNS, and by testing an existing data portability collaboration project: the Data Transfer Project.

7.1 Is there a common definition for a social networking site?

The following definition was adopted for the thesis.
A social networking site is a networked communication platform where:

- *users have unique profiles that consist of content provided by the user;*
- *users can publicly articulate connections that can be viewed and roamed by others;*
- *users can consume, produce, and/or interact with streams of content generated by users that are part of their connections on the SNS [13].*

As stated in section 2, different considerations and the relatively old age of the articles defining SNSs [6] [13] make it difficult to construct a proper and definitive definition of an SNS. More research into this subject might lead to new insights and considerations to possibly construct a more definitive definition of an SNS.

7.2 Was data portability facilitated?

Of the 21 examined SNSs, 10 offered automated personal data downloads in a machine-readable format as described in this thesis, which is 47.62%. From these 9 SNSs, 2 offered personal data uploads in a machine-readable format as described in this thesis, which would be 2 out of 21, 9.52%. Nearly half of the sites facilitated an automated method of personal data download. While it is not a hard requirement, it does make it more user-friendly. Sites that did not have the download option readily available, were willing to make an export of the personal data for the download when requested by mail or any other messaging system. This method does take more steps to achieve the goal, but it is correct in the way the right to data portability is described.

Direct data transfer options were available on Facebook and, therefore, was examined in this regard. Where other sites offer no direct data transfer at the time of writing this thesis, Facebook offers direct data transfer options to ten target sites already. For other sites it might not be technically feasible yet to implement direct data tranfer.

7.3 How was data portability facilitated?

File download

File download was facilitated on the majority of examined sites. Table 3 shows that this was done mostly in HTML and JSON format, and some in XML, CSV, TXT or JPG. In some cases, the JPG would be added in a zip file containing the HTML or JSON file(s). Usually the provided ZIP file would contain one or more data files with the entire collection of personal data. As some sites offered only HTML or only TXT in the personal data download, it does not fully comply with the right to data portability, since HTML is not considered a machine-readable file. It would seem that article 20 of the GDPR is difficult carry out in this regard.

As stated previously in the thesis, next to a larger JSON file download for posts and personalia, Mastodon allows for different subsections to download specific CSV files containing only a subset of data, such as a list of followed accounts.

File upload

Mastodon offered file upload for the CSV files that they offer as downloads, which worked well in the upload test.

Micro.blog offered XML file upload for posts, but this only worked with the XML downloaded from Micro.blog. All other uploads did not work as expected and resulted in no change on the site (table 5).

With both Mastodon and Micro.blog, their own data downloads were able to be successfully uploaded. This is not a surprising feature, since a site would not implement different data models for itself. With cross-platform testing on Micro.blog, it is unfortunate that uploads do not work, as that would be real world scenario we are trying to explore. Data standardization may be key to improving this scenario [16], where in this case one or more shared data schemas would be most beneficial. Unfortunately, in all downloaded files there are differences in the naming conventions within the schema. This shows that it is difficult for different companies to coordinate and adjust schemas with each other, and is expected to be even more difficult for competitors to accomplish.

Direct transfer

Facebook offers direct data transfer to ten target sites, but it is worth noting that of these 10 target sites, none is a comparable service to Facebook, they are not alternatives to the Facebook site. It was initially expected that transfer would be made possible to comparable services, perhaps even competitors.

Direct transfer did not always work as expected. Folder structure was lost when sending photos to Google Photos in the first test, and all photos were transferred when selecting a specific date range photos. Photos should also be placed in the expected location, as in some cases all photos were placed in the main directory of user on the target site. On other sites, based on the type of data, transfer would be done into a new folder when storage sites were the targets, and posts would appear are as posts on target sites specifically for posts. Events would not be transferred to Google Calendar.

Facebook has put in a considerable amount of work to make this amount of direct transfer possiblities. But since there is no case where complete data sets are transferred, but only subsets of data, this is inconsistent with the data download option. It might be considered more appropriate to adjust the GDPR for this discrepancy and to consider which method would be more beneficial to both user (data subject) and site (data controller).

OAuth data retrieval

Only Sociall offered OAuth data retrieval, but unfortunately contained an error preventing the data retrieval test. Therefore, no sites offered any operational OAuth data retrieval functionality.

7.4 What was preferred by the users?

Figure 5 in section 5.2 displays direct data transfer from a target site as least preferable (16.13%). Using OAuth from the target site to retrieve data from a source site seems to have the largest preference (45.16%). It is worth noting that downloading and uploading a machine-readable file was more preferable (32.26%) than the first mentioned direct transfer due to control over data and safety concerns. Therefore, it can be stated that users do have preferences, and the majority either wants to use OAuth or download and upload a file.

Transfer errors

When mockup experiment participants were faced with data transfer errors and were made aware that they would be asked about it in the survey, 4 out of 10 new respondents noticed the transfer error where posts were transferred, while only personalia and no posts should have been transferred. 6 out of 10 did not notice the error. As more respondents noticed the error, more were not likely to use direct data transfer. While nobody likes transfer errors, it is not a surprise that people are less likely use a transfer method containing bugs. However, since the errors in the mockup experiment are based on actuak errors with direct transfer on live sites, it shows that data transfer has not matured yet since the GDPR came into effect in 2018.

7.5 Compliance with the right to data portability

When sites offer automated personal data downloads in a machine-readable format, it is usually contained in a ZIP file containing all provided personal data. This is compliant with article 20 of the GDPR (right to data portability). Having data transmitted directly from one controller to another, only offered by Facebook at the time of writing this thesis, only subsets of personal data (only posts or only images and videos) can be transferred. This data is not equal to the personal data provided in a personal data download ZIP file, nor is it sent to a site with equal functionality. It would seem that a subset of personal data is compliant to the GDPR as the right to data portability does not specify which personal data, all or a subset, should be provided to the data subject or transmitted to another data controller [1].

7.6 Adjust approach to data portability

With the approach of providing or transmitting data subsets also being GDPR compliant if it entails the request of the user, a different approach than the majority of current implementations might be suggested.

In the case of personal data downloads, Mastodon demonstrates a working download and upload function for subsets of data in CSV format. Keeping personal data files smaller and making it easier to manage which data the user would want to receive and possibly reuse. This would make it more user-friendly and might make users more prone to using such features.

With direct data transfer, Facebook offers data transfer to 10 target sites, all of which are not equal to Facebook, but share parts of functionality, such as photos/videos which can be placed on both Facebook and Google Photos. Facebook has put in a considerable amount of work in comparison to other sites, since it is the only one offering such an option as of now, making data transfer easier for its users. Linking with other major sites such as Google Photos, Dropbox, etc. also makes it more prone to be used.

Our advice for article 20 of the GDPR and websites would be to focus more on partial overlap between sites (data controllers) and partitioning data in subsets of data a user might provide. This could make the right to data portablity more concrete, allow more control and user-friendliness for the site user (data subject) and better technical feasibility for the website (data controller) that implements data portability. As most survey participants have indicated to prefer data download/upload over direct data transfer and OAuth data retrieval has not been made widely available, the data subset downloads (and/or uploads) could be expected to make the most impact if implemented.

References

1. European parliament and council of the european union: Regulation on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (data protection directive), https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02016R0679-20160504&from=EN
2. Alberini, A., Benhamou, Y.: Data portability and interoperability: an issue that needs to be anticipated in today's it-driven world. Available at SSRN 3038877 (2017)
3. Banda, C.: Enforcing data portability in the context of eu competition law and the gdpr. MIPLC Master Thesis Series (2016/17) (2017)
4. Barth, M.: A case study on data portability. Datenschutz und Datensicherheit-DuD 45(3), 190–197 (2021)
5. Borgogno, O., Colangelo, G.: Data sharing and interoperability: Fostering innovation and competition through apis. Computer Law & Security Review 35(5), 105314 (2019)
6. Boyd, D.M., Ellison, N.B.: Social network sites: Definition, history, and scholarship. Journal of computer-mediated Communication 13(1), 210–230 (2007)
7. Castro, D.: Improving consumer welfare with data portability. Tech. rep., Information Technology and Innovation Foundation (2021)
8. Chen, H.T., Chen, W.: Couldn't or wouldn't? the influence of privacy concerns and self-efficacy in privacy management on privacy protection. Cyberpsychology, Behavior, and Social Networking 18(1), 13–19 (2015)
9. Cravo, D.C.: The right to data portability in eu's gdpr and brazil's lgpd. Brazilian Journal of Law, Technology and Innovation 1(1), 110–140 (2023)
10. De Hert, P., Papakonstantinou, V., Malgieri, G., Beslay, L., Sanchez, I.: The right to data portability in the gdpr: Towards user-centric interoperability of digital services. Computer law & security review 34(2), 193–203 (2018)
11. Diker Vanberg, A., Ünver, M.B.: The right to data portability in the gdpr and eu competition law: odd couple or dynamic duo? European Journal of Law and Technology 8(1) (2017)
12. Egan, E.: Data portability and privacy. White paper, Facebook (September 2019), https://iapp.org/media/pdf/fb_whitepaper_sep_2019.pdf
13. Ellison, N.B., Boyd, D.M.: Sociality through social network sites. In: The Oxford handbook of internet studies (2013)
14. Engels, B.: Data portability among online platforms. Internet Policy Review 5(2) (2016)
15. Facebook, Twitter, Microsoft, Google: Data transfer project overview and fundamentals. Technical overview, Data Transfer Project (July 2018), https://datatransferproject.dev/dtp-overview.pdf
16. Gal, M.S., Rubinfeld, D.L.: Data standardization. NYUL Rev. 94, 737 (2019)
17. Graef, I., Husovec, M., Purtova, N.: Data portability and data control: lessons for an emerging concept in eu law. German Law Journal 19(6), 1359–1398 (2018)
18. Janal, R.: Data portability under the gdpr: A blueprint for access rights? In: Data Access, Consumer Interests and Public Welfare. pp. 319–342. Nomos Verlagsgesellschaft mbH & Co. KG (2021)
19. Kaplan, A.M., Haenlein, M.: Users of the world, unite! the challenges and opportunities of social media. Business horizons 53(1), 59–68 (2010)
20. Krämer, J., Stüdlein, N.: Data portability, data disclosure and data-induced switching costs: Some unintended consequences of the general data protection regulation. Economics Letters 181, 99–103 (2019)
21. Kuebler-Wachendorff, S., Luzsa, R., Kranz, J., Mager, S., Syrmoudis, E., Mayr, S., Grossklags, J.: The right to data portability: Conception, status quo, and future directions. Informatik Spektrum 44, 264–272 (2021)
22. Lehtiniemi, T.: Personal data spaces: An intervention in surveillance capitalism? Surveillance & Society 15(5), 626–639 (2017)
23. Luzsa, R., Mayr, S., Syrmoudis, E., Grossklags, J., Kübler-Wachendorff, S., Kranz, J.: Online service switching intentions and attitudes towards data portability–the role of technology-related attitudes and privacy. In: Proceedings of Mensch und Computer 2022, pp. 1–13 (2022)
24. Martinelli, S.: Sharing data and privacy in the platform economy: the right to data portability and "porting rights". Regulating New Technologies in Uncertain Times pp. 133–152 (2019)
25. Reja, U., Manfreda, K.L., Hlebec, V., Vehovar, V.: Open-ended vs. close-ended questions in web questionnaires. Developments in applied statistics 19(1), 159–177 (2003)
26. Sørum, H., Presthus, W.: Dude, where's my data? the gdpr in practice, from a consumer's point of view. Information Technology & People (2020)
27. Syrmoudis, E., Mager, S., Kuebler-Wachendorff, S., Pizzinini, P., Grossklags, J., Kranz, J.: Data portability between online services: An empirical analysis on the effectiveness of gdpr art. 20. Proceedings on Privacy Enhancing Technologies 3, 351–372 (2021)

28. Ursic, H.: Unfolding the new-born right to data portability: Four gateways to data subject control. SCRIPTed 15, 42 (2018)
29. Vahedi, Z., Zannella, L.: The association between self-reported depressive symptoms and the use of social networking sites (sns): A meta-analysis. Current Psychology pp. 1–16 (2019)
30. Wong, J., Henderson, T.: The right to data portability in practice: exploring the implications of the technologically neutral gdpr. International Data Privacy Law 9(3), 173–191 (2019)

Appendices

A File upload

A.1 Mastodon

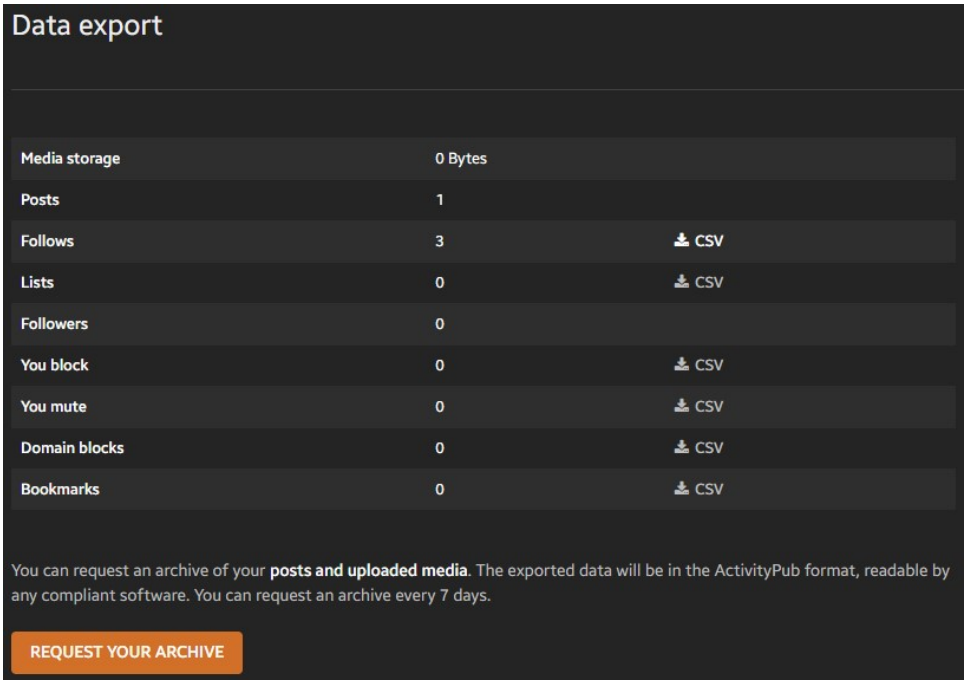

Fig. 9. List of possible Mastodon (donphan.social) CSV exports and archive request - Requires more clicks to get all data

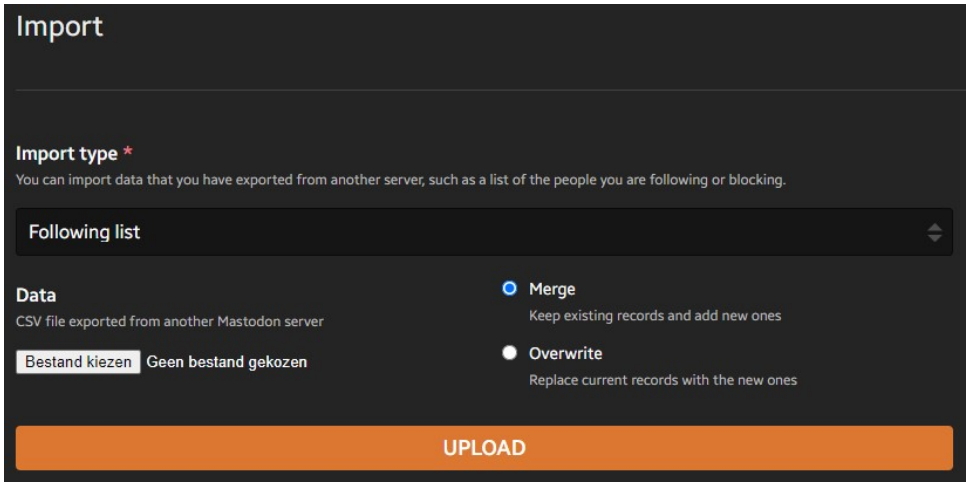

Fig. 10. Mastodon (donphan.social) import section - Works well with files in the format provided in the Mastodon export format

A.2 Micro.blog

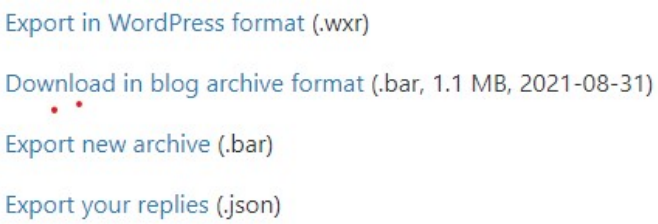

Fig. 11. Micro.blog export section

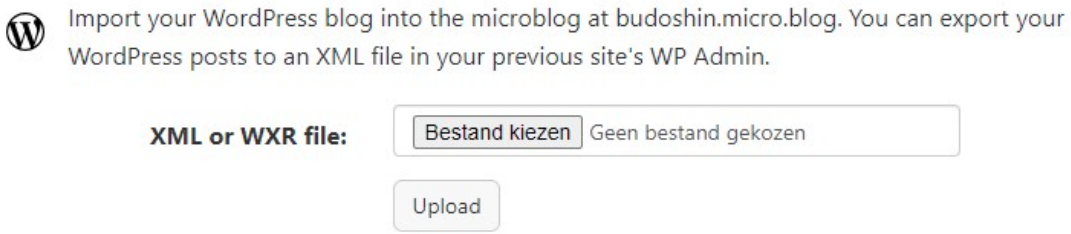

Fig. 12. Micro.blog WXR and XML import section

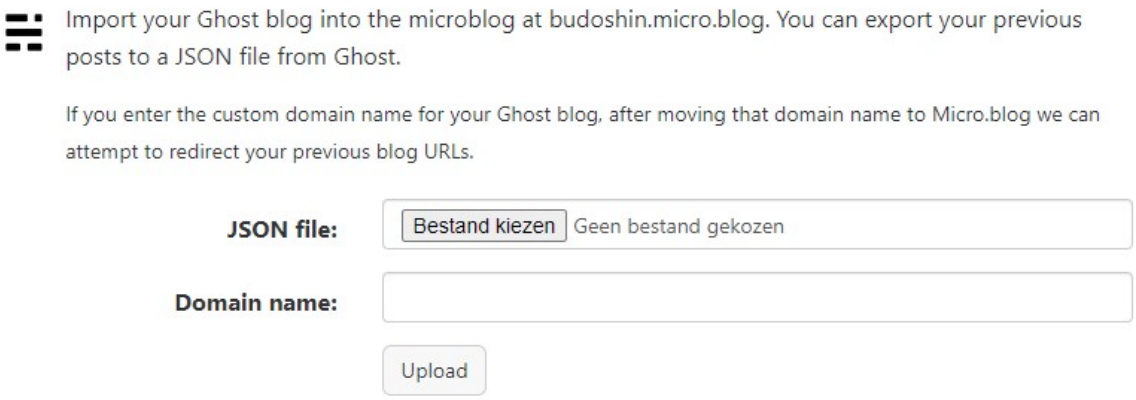

Fig. 13. Micro.blog Ghost JSON import section

Fig. 14. Micro.blog Foursquare/Swarm import section

B Mockup experiment

B.1 Indirect data transfer

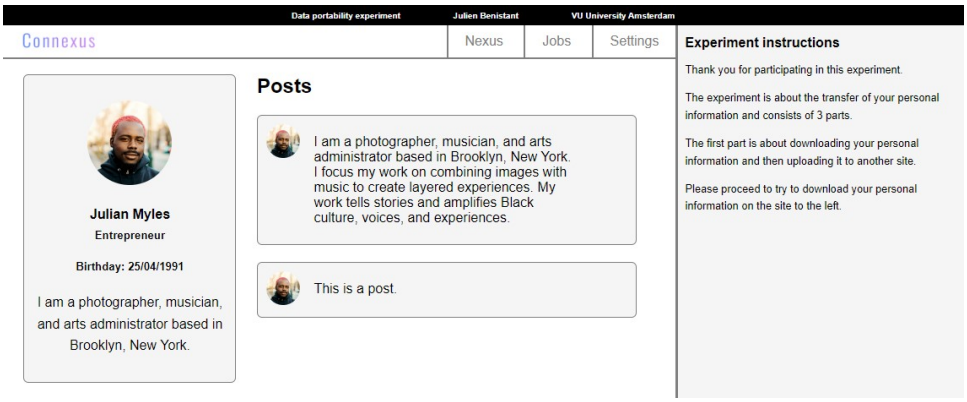

Fig. 15. Mockup experiment - Indirect transfer - Part 1 Step 1 - Landing page
The participant sees the profile of the user, which has all fields filled, such as profile picture, name, job title, birthday, short description, and has two posts. The first instructions are shown on the right side of the screen and the user has to find the option by themselves to download their personal data.

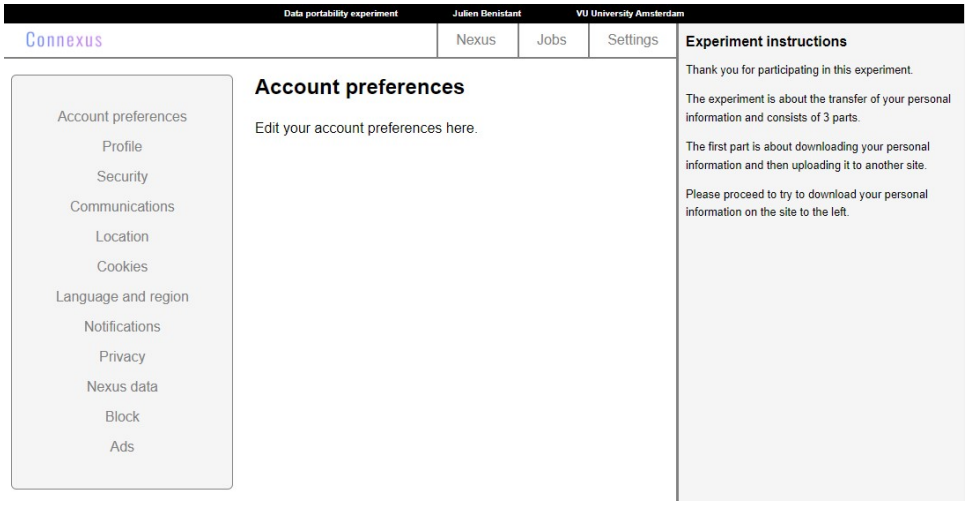

Fig. 16. Mockup experiment - Indirect transfer - Part 1 Step 1 - Settings
After user has clicked the settings button, a submenu appears on the left side of the screen with twelve options. All options are clickable, but only one leads to the page with the download option.

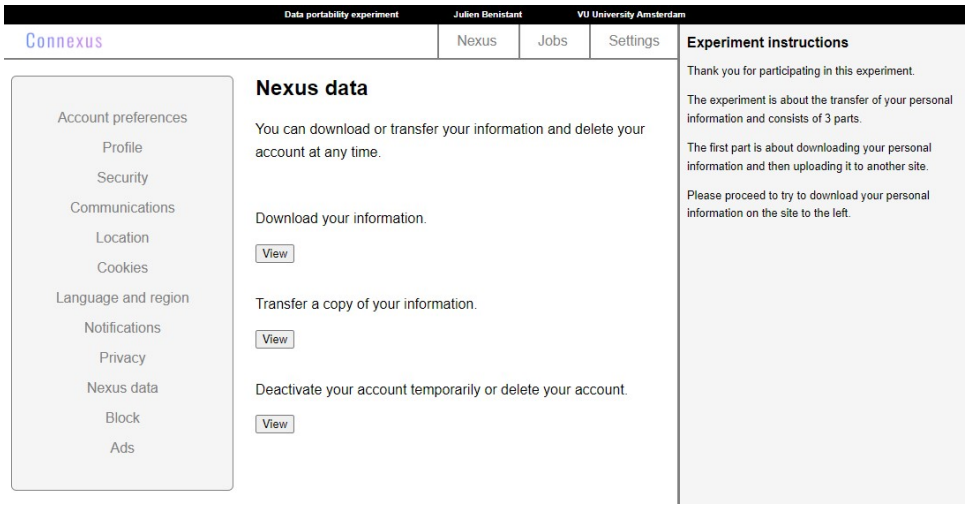

Fig. 17. Mockup experiment - Indirect transfer - Part 1 Step 1 - Nexus sata
The Nexus data screen holds the different options in regard to personal data. Only the View button beneath "Download your information" is a link to a different page to prevent the user from having to user the "back/previous" button of the browser.

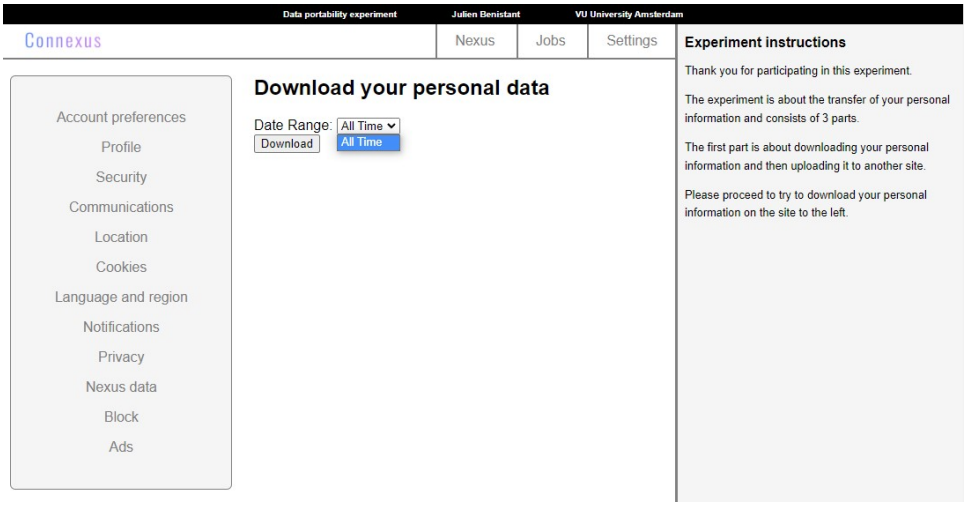

Fig. 18. Mockup experiment - Indirect transfer - Part 1 Step 1 - Download page
The download page shows a dropdown list called "Date range" to have a similar feel to live websites with this functionality. Only one option is available for the user and the only thing to do further is press Download.

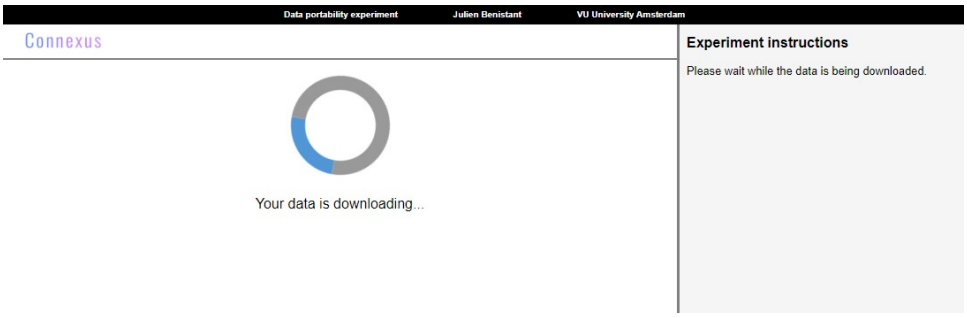

Fig. 19. Mockup experiment - Indirect transfer - Part 1 Step 1 - Data download
Loading screen with a set timer to simulate a download in progress. On live sites, it will take a longer period of time for the data to be ready for download.

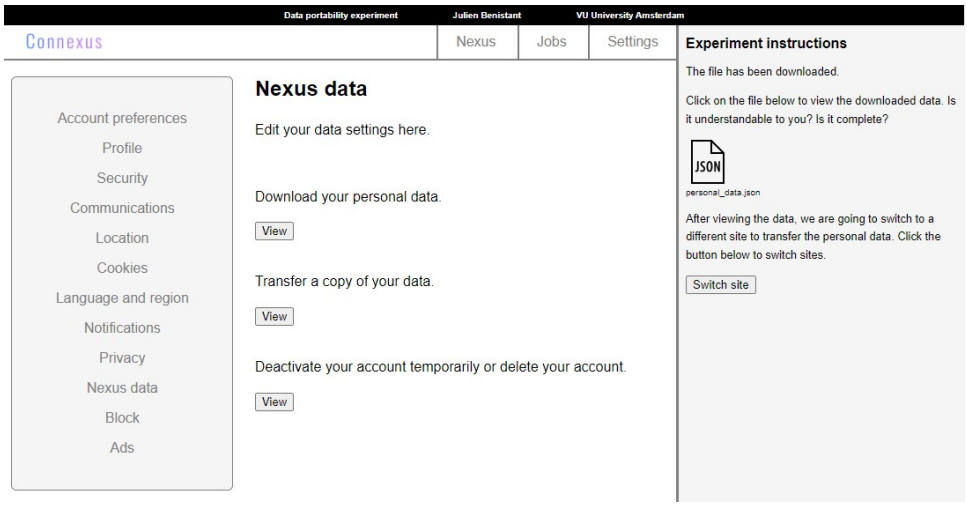

Fig. 20. Mockup experiment - Indirect transfer - Part 1 Step 2 - Click file to open
After the download screen, the "downloaded" file appears in the instruction screen as a clickable icon. The instruction tells the participant to click on the file and view the downloaded data.

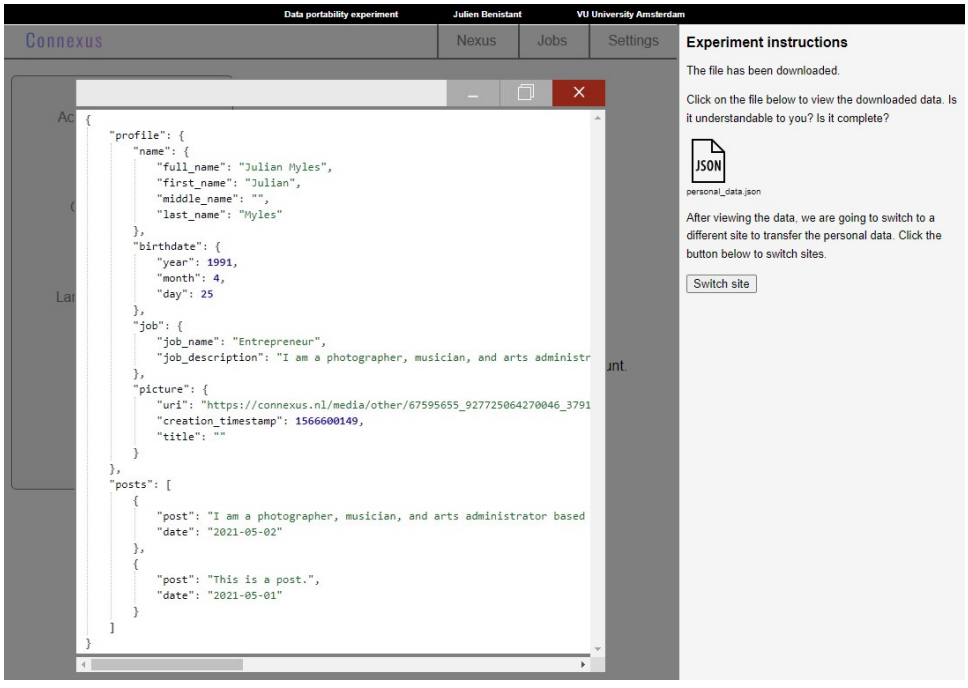

Fig. 21. Mockup experiment - Indirect transfer - Part 1 Step 2 - View data and switch site button
A pop up screen is shown when the file icon is clicked, containing the personal data of the user in JSON format.

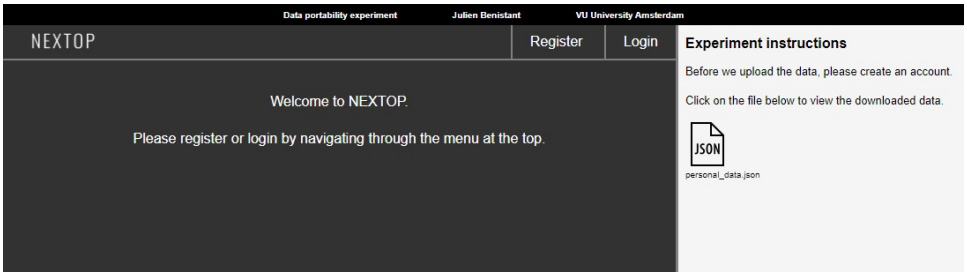

Fig. 22. Mockup experiment - Indirect transfer - Part 1 Step 3 - NEXTOP landing page
This is the landing page of the fictional target site where there is no user logged in. The instruction asks the user to create an account. It is also still possible to view the downloaded JSON data.

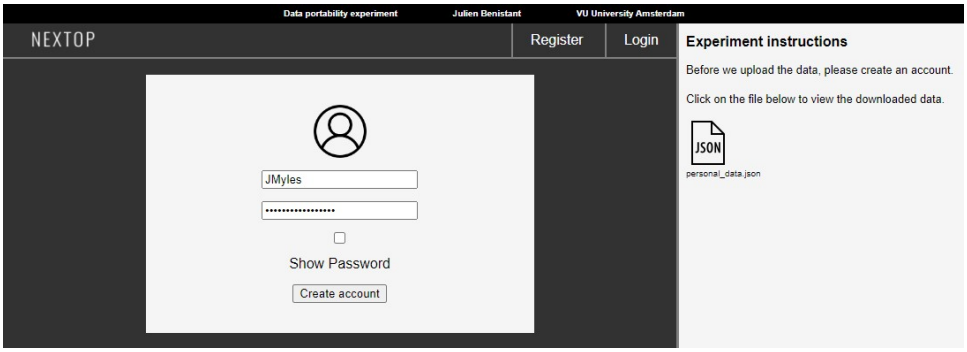

Fig. 23. Mockup experiment - Indirect transfer - Part 1 Step 3 - Register
Here, the registration page is shown with minimal input requirements, a username and a password, which are already filled in for consistency with the user account pages. The user can directly click on the Create account button.

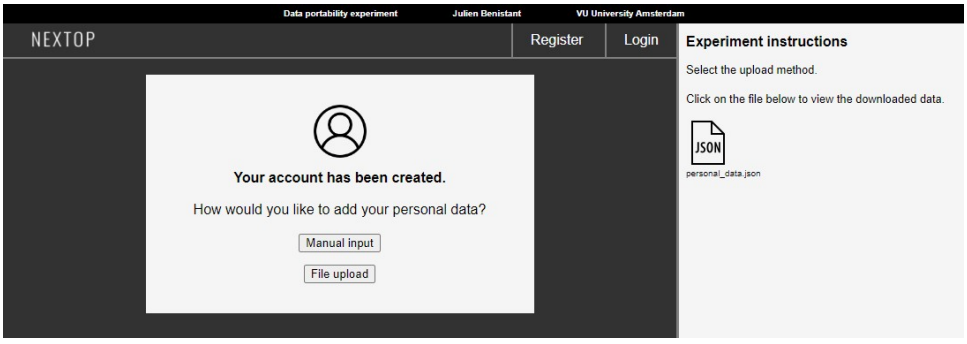

Fig. 24. Mockup experiment - Indirect transfer - Part 1 Step 4 - Select file upload
Now that the user has a new account which can receive data, the user has to select whether to use manual input or file upload. Only the latter leads to the next step of the experiment. The former does nothing.

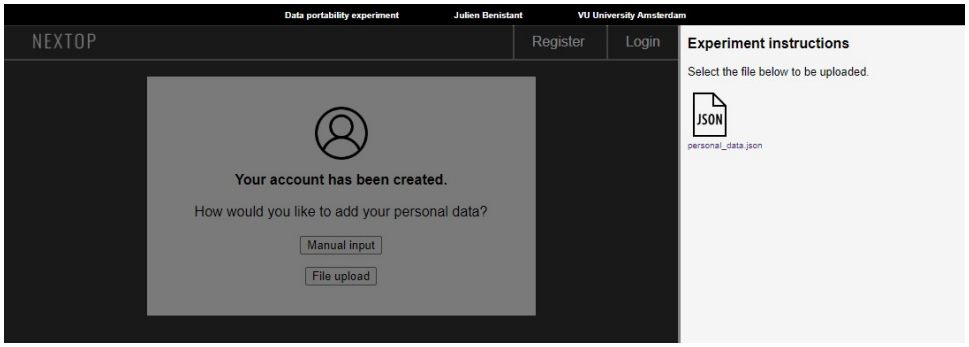

Fig. 25. Mockup experiment - Indirect transfer - Part 1 Step 5 - Click file
The user is prompted to "select" the file to be uploaded from the instruction screen. This is done by clicking on the file icon.

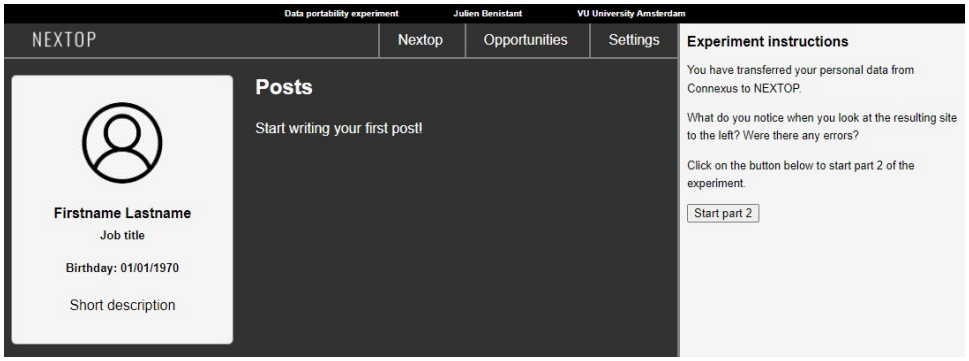

Fig. 26. Mockup experiment - Indirect transfer - Part 1 Step 6 - Upload complete
This is the resulting page after de personal data has been uploaded. We see here that everything is missing: profile picture, name, job title, birthday, short description, and the two posts, as shown on the first screen of the experiment.

B.2 Direct data transfer

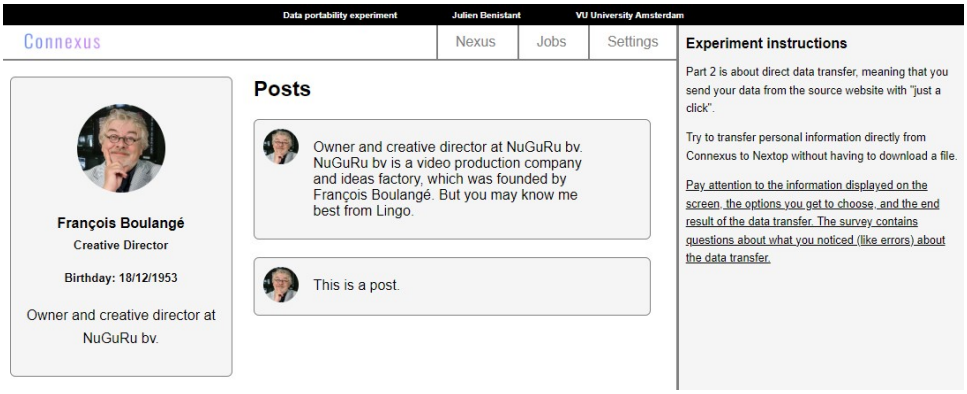

Fig. 27. Mockup experiment - Direct transfer - Part 2 Step 1 - Landing page
The participant sees the profile of the user, which has all fields filled, such as profile picture, name, job title, birthday, short description, and has two posts. Both the name and a post contain special characters, The first instructions are shown on the right side of the screen and the user has to find the option by themselves to transfer their personal data.

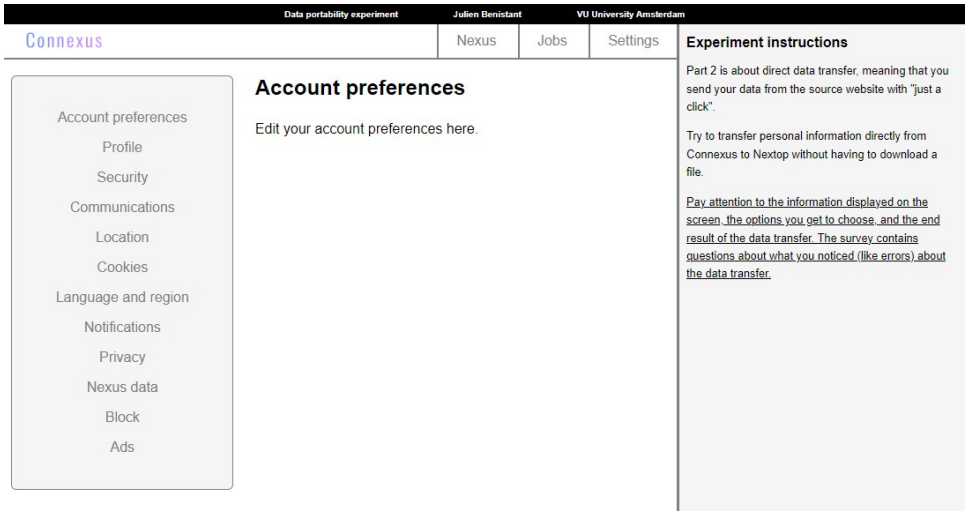

Fig. 28. Mockup experiment - Direct transfer - Part 2 Step 1 - Settings
After the user has clicked the settings button, a submenu appears on the left side of the screen with twelve options. All options are clickable, but only one leads to the page with the data transfer option.

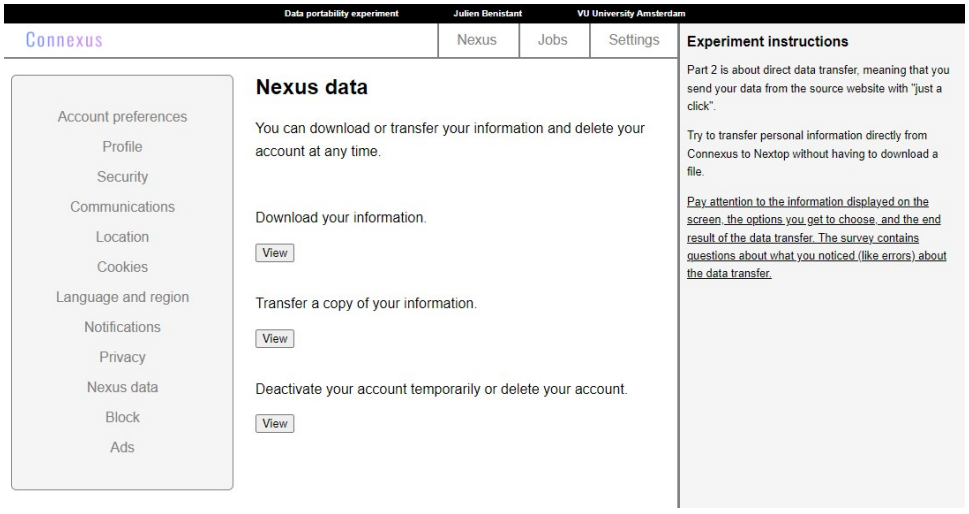

Fig. 29. Mockup experiment - Direct transfer - Part 2 Step 1 - Nexus data - Transfer a copy
The Nexus data screen holds the different options in regard to personal data. Only the View button beneath "Transfer a copy of your information" is a link to a different page to prevent the user from having to user the "back/previous" button of the browser.

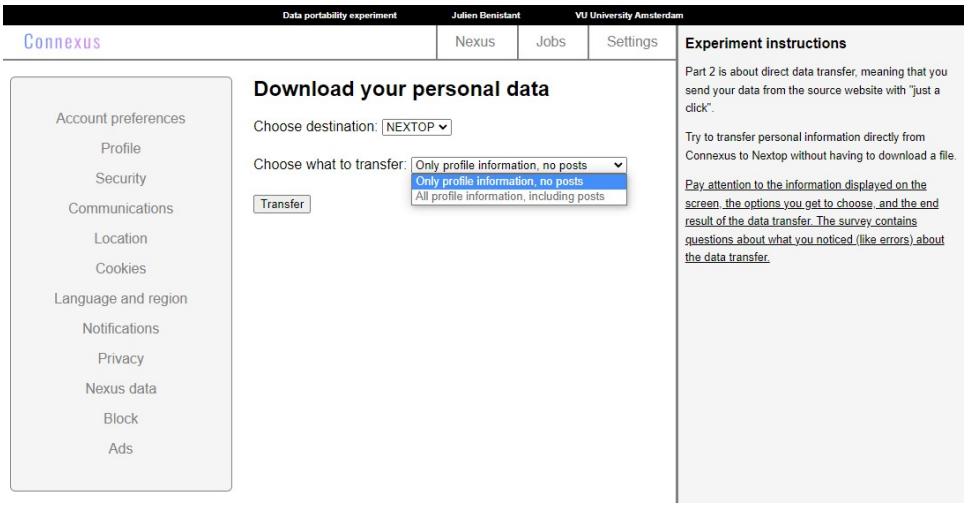

Fig. 30. Mockup experiment - Direct transfer - Part 2 Step 1 - Data transfer page
The download page shows a dropdown list for destination, i.e. the target site, and a dropdown list with options of what type(s) of data to transfer. Only one option is available for the user for both options. The goal of the dropdown lists is to inspect which option is supposed to be enacted. The only thing to do further is to press Transfer.

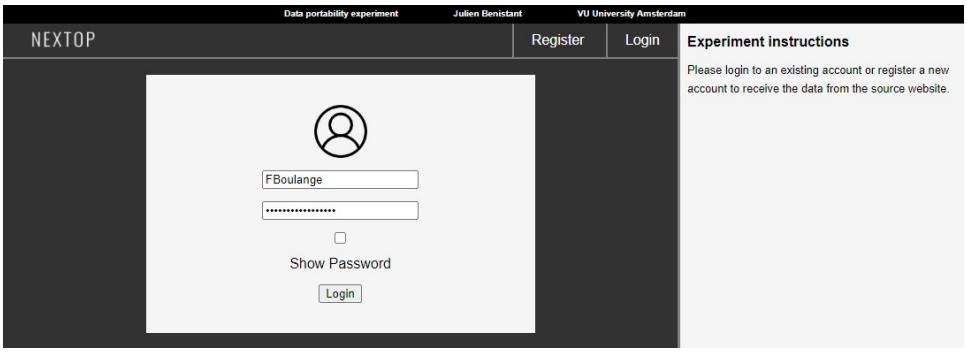

Fig. 31. Mockup experiment - Direct transfer - Part 2 Step 2 - NEXTOP login
In this example, the user is asked to register or login. Only the login option is available with prefilled fields for consistency.

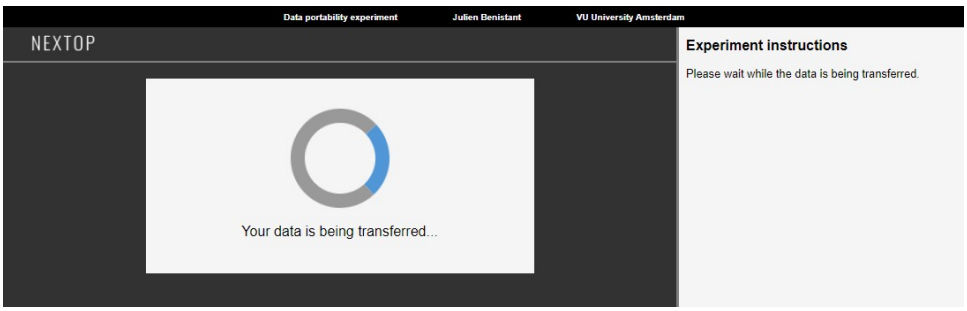

Fig. 32. Mockup experiment - Direct transfer - Part 2 Step 2 - Data transfer
Loading screen with a set timer to simulate a data transfer in progress. On live sites, it might take a longer period of time for the data to be transferred.

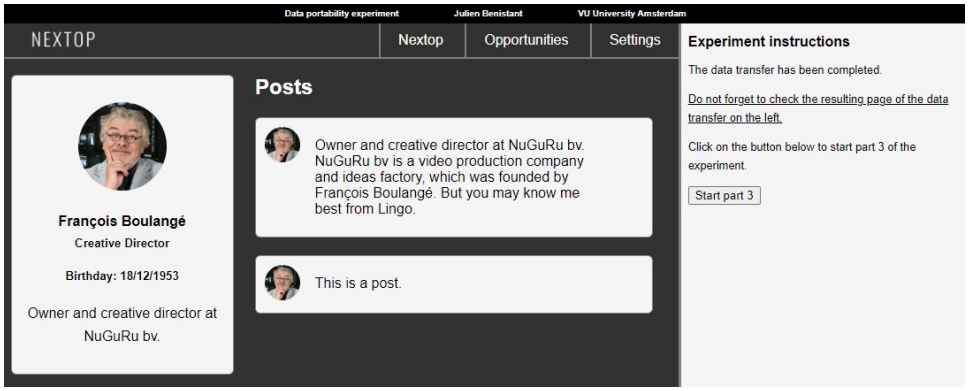

Fig. 33. Mockup experiment - Direct transfer - Part 2 Step 3 - Resulting target page
The resulting page after the transfer shows no direct signs of faulty transfer. However, the mistake lies in the selected options, where only profile information was supposed to be transferred (see figure 30), the posts were also transferred, as mentioned in section 4.2 and analyzed in section 5.2. This is based on an occurrence on a live site with image transfer.

B.3 Data transfer by OAuth

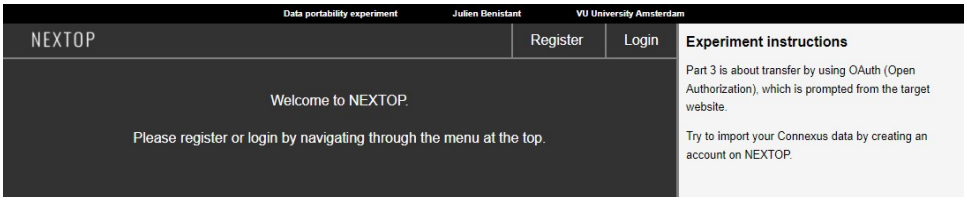

Fig. 34. Mockup experiment - OAuth - Part 3 Step 1 - Start on target site
The user is requested to create an account on the target site.

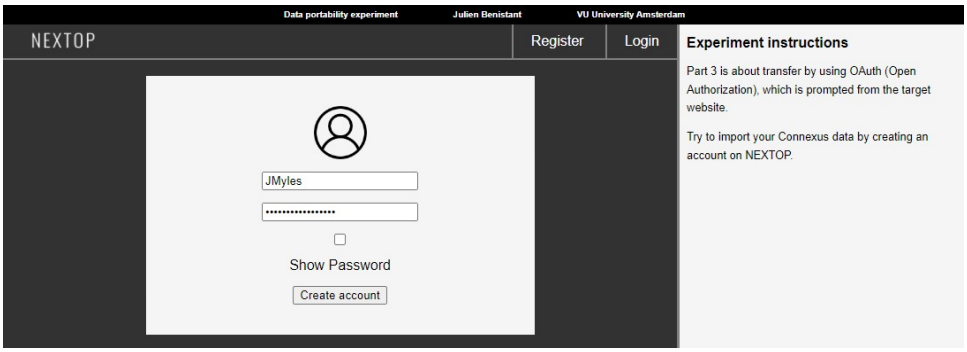

Fig. 35. Mockup experiment - OAuth - Part 3 Step 1 - Register
On the registration page, prefilled fields are shown for consistency. The user can only move to the next part by clicking Create account.

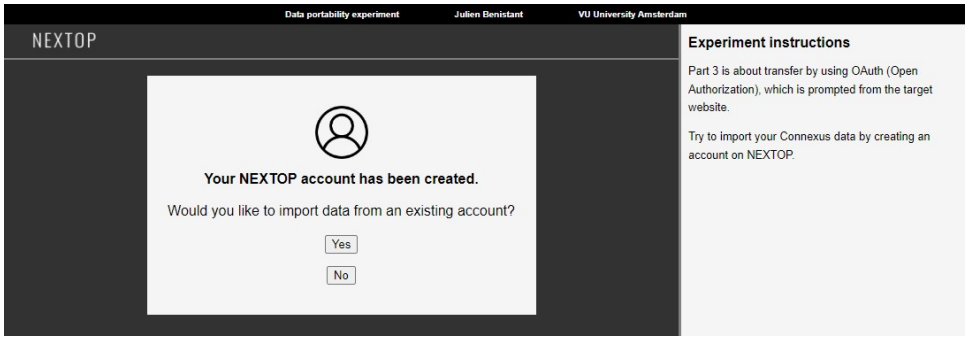

Fig. 36. Mockup experiment - OAuth - Part 3 Step 1 - Data import options
When the account is created, the user is asked if data should be imported from an existing account. The instructions explain that this entails an existing account on the source site. Yes is the only button that leads to the next step.

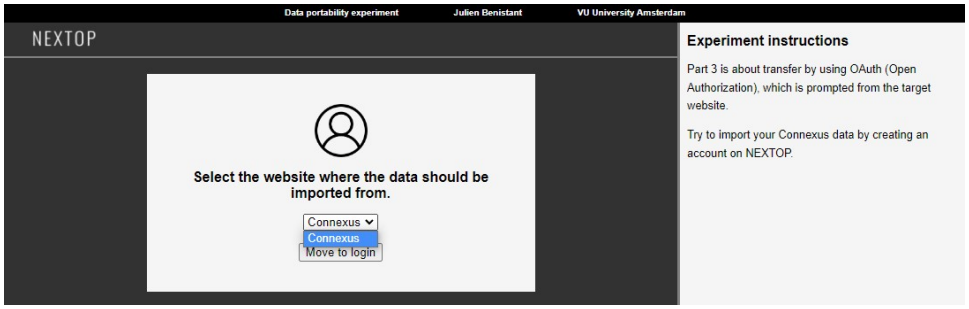

Fig. 37. Mockup experiment - OAuth - Part 3 Step 2 - Select source site
A dropdown list is shown for accuracy with regard to live sites, but only shows the one option that we are interested in. Next, the user has to click the Move to login button.

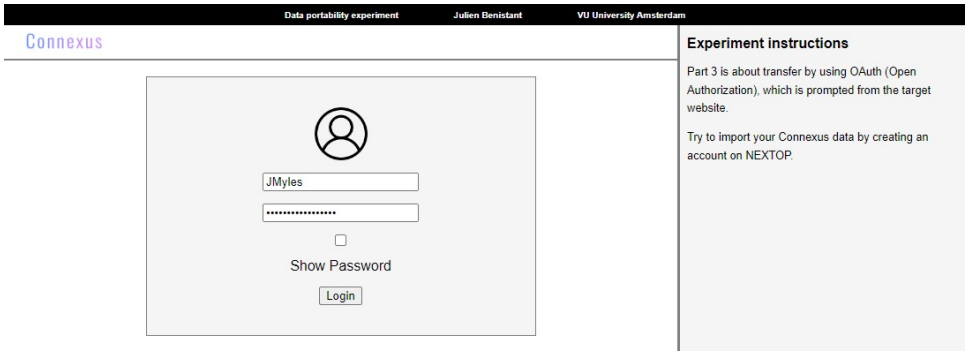

Fig. 38. Mockup experiment - OAuth - Part 3 Step 2 - Login to source site
The user lands on the login screen of the source site in order to gain access to their personal data.

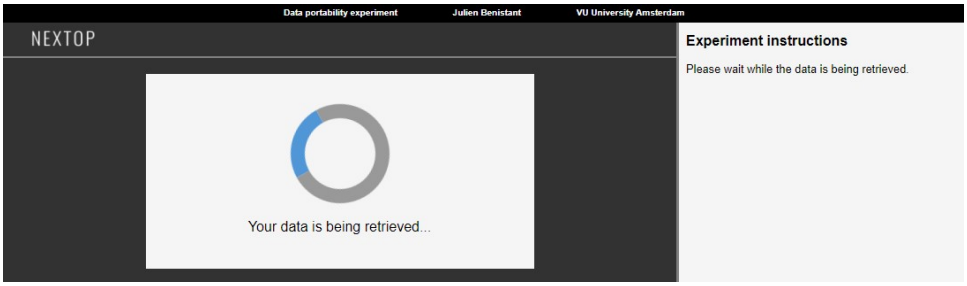

Fig. 39. Mockup experiment - OAuth - Part 3 Step 2 - Data retrieval
After logging in, the data retrieval starts. Data retrieval is shown as a loading screen with a set timer to simulate a data transfer in progress. On live sites, it might take a longer period of time for the data to be transferred.

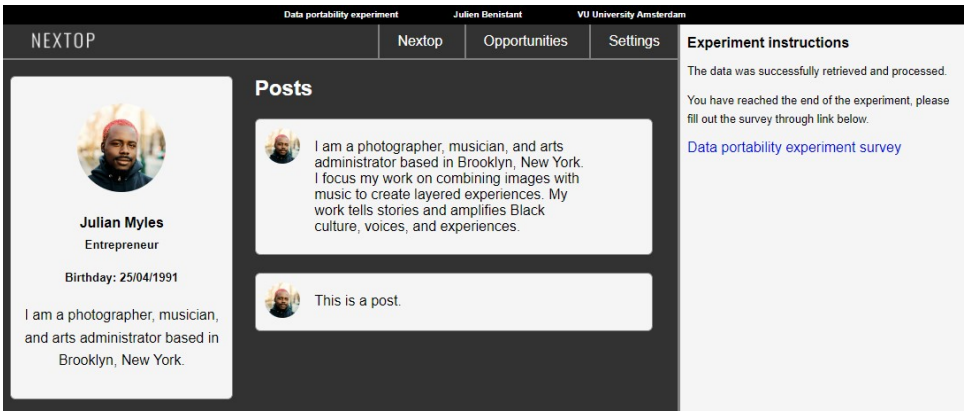

Fig. 40. Mockup experiment - OAuth - Part 3 Step 3 - Data retrieval complete
This page shows the resulting profile with the personal data from the source site. In the instructions a link appears for the participant to start the survey of the performed experiments.

C Mockup experiment survey

Placeholder for embedding

Demographic information

Thank you for participating in the experiment. This survey contains 25 questions spread over 4 pages.

At the end of the survey, it is optional to enter your e-mail address for a chance to win an Amazon gift card. Your e-mail address will not be shared or sold to any other party, or be used for any other purpose than to determine a winner who will receive the gift card.

We wish to examine whether social networking site users are aware of their data portability rights and options. First, we would like to investigate if demographic information and several general questions have any significance for the results.

What is your current age?

- ◯ Under 18
- ◯ 18 - 24
- ◯ 25 - 34
- ◯ 35 - 44
- ◯ 45 - 54
- ◯ 55 - 64
- ◯ 65 or over

What is your gender?

- ◯ Male
- ◯ Female
- ◯ Other
- ◯ I do not wish to share this information

What is the highest level of education you have been enrolled in?

- ◯ No schooling completed
- ◯ Nursery school to 8th grade
- ◯ Some high school, no diploma
- ◯ High school graduate, diploma or the equivalent (for example: GED)
- ◯ Some college credit, no degree
- ◯ Trade/technical/vocational/middle-management training
- ◯ Associate degree
- ◯ Bachelor’s degree
- ◯ Master’s degree
- ◯ Doctorate degree

Do you have any technical expertise or background? If so, what kind of expertise/background?
For example: You work with information technology or follow an IT-related study.

The General Data Protection Regulation (GDPR) is the leading regulation in the European Union concerning privacy rights of consumers. Its purpose is to protect personal data and strengthen the rights of civilians in a society which is becoming increasingly more digital.
Source: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX\%3A02016R0679-20160504\#tocId7

Have you previously heard about the GDPR? If so, how and what did you learn about it?

What types of data would you consider to be part of your personal/self-provided data on social networking sites? Examples of types of data: personal (such as name and age), posts, connections, etc.

Indirect data tranfer

The following questions are about part 1 of the experiment. In this part, a personal data file was downloaded from one site and uploaded to another site.

Have you ever retrieved or transferred your personal data from/to a social networking site by downloading and uploading a data file and why? If not, were you aware that such functionality exists or should exist?

When retrieving personal data from a social networking site, the data should be provided in a machine-readable format. Examples of machine-readable file extensions are XML, JSON, and CSV.

Which of the mentioned formats are you familiar with?

- [ ] XML
- [ ] JSON
- [ ] CSV
- [ ] Other, namely...
- [ ] None of the above

During the experiment, a personal data file was downloaded. The file was in JSON format. Was the data shown in

the file understandable to you?

As you viewed the data in the JSON file, was there any personal information of the user that you missed that was present on the source site Connexus? If yes, could you tell us what you missed?

When the personal data file was uploaded to target site NEXTOP, the data did not appear on the new site in the experiment. What might be the reason for this?

If you were to switch social networking sites, i.e. remove your account on one and create it on another, would you consider this method of downloading and uploading a file? Why?

Direct data transfer

The following questions are about part 2 of the experiment. This part of the experiment had you send your information from the source site, Connexus, to the target site, NEXTOP, by clicking a "Transfer data" button. This functionality is built and executed by the source site.

Have you ever encountered this method of data transfer? If yes, where?

Direct data transfer offers ease of data transfer to the website users when compared to indirect data transfer in part 1. If this option would be readily available on all social sites, how likely would you be to use the functionality to switch platforms? And why?

The data transfer had a different result than expected. Did you notice anything wrong the transfer? If so, what did you notice?

In the experiment, a page was shown on the source site with options on what to transfer. The only option you can select is to transfer personal information and no posts. The result of the transfer, however, shows a page where

both personal information and posts were transferred.

Would such errors change your answer about te consideration of using this method of data transfer? Why?

Data transfer using OAuth

The following questions are about part 3 of the experiment and its relation with the first two parts. This part had you retrieve your information from the source site, Connexus, as you attempted to create an account on the target site, NEXTOP.

The retrieval of the information was done using OAuth. OAuth stands for *Open Authentication* and is a technological standard that allows you to share information between services without exposing your password. An example of OAuth would be to log in to sites using your Google or Facebook account.
Source: https://www.hp.com/us-en/shop/tech-takes/what-is-oauth

As you have gone through the steps of retrieving your data using OAuth, how likely would you be to use this functionality yourself and why?

This part of the experiment did not demonstrate any errors. However, errors might occur on actual sites using this method. For example, the method might simply give an error message if some form of developer license is missing. Another example might be missing data or data in wrong places.

How likely would you be to use the OAuth method of data transfer?

If you were to switch from one social networking site to another and use a data portability option for the transfer of your personal data, which would have your preference?

- ◯ Download and upload a machine-readable file
- ◯ Direct data transfer from the site you are leaving
- ◯ OAuth from the site you are joining to directly retrieve data from the site that you are leaving
- ◯ No preference

Please elaborate on your choice made in the previous question.

Ease of understanding: Please select to what degree you agree or disagree with the following statements.

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
Indirect data transfer (download and upload) is easy to understand	○	○	○	○	○
Direct data transfer (select options on source site to send to target site) is easy to understand	○	○	○	○	○
OAuth data transfer (target site downloads from source site) is easy to understand	○	○	○	○	○
The data transfer possibilities are easy to distinguish from each other	○	○	○	○	○
It is clear on real websites which possibilities are available	○	○	○	○	○

Usability: Please select to what degree you agree or disagree with the following statements.

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
Indirect data transfer (download and upload) is easy to use	○	○	○	○	○
Direct data transfer (select options on source site to send to target site) is easy to use	○	○	○	○	○
OAuth data transfer (target site downloads from source site) is easy to use	○	○	○	○	○

The right to data portability is covered in article 20 of the General Data Protection Regulation (GDPR), known in the Netherlands as the Algemene verordening gegevensbescherming (AVG).

The right to data portability gives users the right to retrieve the personal data that they have provided to a controller (such as a social networking site) and/or have the controller send the data to a different controller. They are also obliged to develop functionality to transfer your data for you where it is feasible for them.

Before this experiment, were you aware of the right to data portability?

- ○ Yes
- ○ No

If you answered yes on the previous question, how did you become aware?

Optional: Chance to win gift card

Optional: Enter your e-mail address for a chance to win a gift card of your choice worth €50,-.