HOW PORTABLE IS PORTABLE?
The GDPR’s right to data portability in practice

Janis Wong and Tristan Henderson

School of Computer Science
University of St Andrews
jccw@st-andrews.ac.uk
GDPR?

https://www.thephoenix.ie/2017/07/cartoon-3515-10/
DATA PROTECTION

- Dutch population registers from 1849
  - Name, Birth, Religion, Marital Status…
- Wide-scale misuse in WWII
- Do we have rights over our personal data?
  - Balancing our rights with potential misuse

https://www.dutchgenealogy.nl/population-registers/
• New data protection regulation into force from May 2018
• Strengthens existing data subject rights (e.g. access, erasure, rectification)
• Introduces one new right: the right to data portability
• Data subject rights now free to exercise
• Aims to protect the processing of personal data of natural persons under a “technologically neutral” framework
• **Personal data:** “any information concerning an identified or identifiable natural person” (GDPR Recital 26)

• **Data subject:** those about whom personal data are collected and are about
  - The GDPR strengthens and introduces data subject rights

• **Data controller:** those who collect or determine what these data are used for
  - The GDPR enforces legal obligations on data controllers

• **Data processor:** those who process data on behalf of the data controller

• Defined by the ICO as the UK Data Protection Authority
  - https://ico.org.uk/
DATA SUBJECT RIGHTS UNDER THE GDPR

- Right to be informed (Article 12)
- Right of access (Articles 12, 15)
- Right to rectification (Article 16)
- Right to erasure ‘right to be forgotten’ (Article 17)
- Right to restriction of processing (Article 18)
- Right to data portability (Article 20)
- Right to object (Article 21)
- Rights related to automated decision-making (Article 22)
DATA SUBJECT RIGHTS UNDER THE GDPR

- Right to be informed (Article 12)
- Right of access (Articles 12, 15)
- Right to rectification (Article 16)
- Right to erasure ‘right to be forgotten’ (Article 17)
- Right to restriction of processing (Article 18)
- Right to data portability (Article 20) (NEW!)
- Right to object (Article 21)
- Rights related to automated decision-making (Article 22)
• Now that the GDPR is here:
  • How do data controllers approach Article 20 (RtDP)?
  • Do they know how to meet their RtDP obligations?
• How do we answer this?
  • Make some requests and look at the responses
Right to data portability

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:

1.1 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

1.2 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.
HOW DO DATA CONTROLLERS KNOW HOW TO COMPLY?

• The ICO issued the Guide to the GDPR\(^1\), including:
  • Definitions to key terms
  • Suggested data formats
  • Data controllers’ obligations to data subjects

• The A29WP (now EDPB) published Guidelines on the right to data portability\(^2\), clarifying:
  • What data should be included in responses to RtDP requests
  • How portable data must be provided
  • Interoperability and interoperable systems only as desired outcomes


\(^2\) ec.europa.eu/newsroom/document.cfm?doc_id=44099
- Created a Python program to automatically generate and send RtDP requests
- Used e-mail since this is the only universally supported mechanism for making requests
- Sent 230 requests to a variety of data controllers
- Study ran from 25 May 2018 to 26 August 2018
The RtDP process simplified

- E-mail RtDP Request for Personal Data Used on Social Network
- E-mail RtDP Response with Personal Data in a portable file format
WHAT KIND OF DATA CONTROLLERS?

A wide variety of data controllers — not just SNSes!
172 of 230 data controllers responded, with 85% responding within a month.
The wide range of response formats represent different data types, with no consensus on what is most appropriate.
WHAT FILE FORMATS DID CONTROLLERS USE?

The wide range of response formats represent different data types, with no consensus on what is most appropriate.

Categorised File Formats from Data Controller Responses

Lots of commonly used formats (CSV, TXT, DOC), but were they structured and machine-readable?
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Email body</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>CSV</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>DOC/DOCX</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>EML</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>HTML</td>
<td>✂</td>
<td>✓</td>
<td>✂</td>
</tr>
<tr>
<td>ICS</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>JPEG</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>JSON</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>KMZ</td>
<td>✓</td>
<td>x</td>
<td>✓</td>
</tr>
<tr>
<td>MBOX</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>MP4</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>PDF</td>
<td>✂</td>
<td>✓</td>
<td>✂</td>
</tr>
<tr>
<td>PNG</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>RTF</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>TEX</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>TXT</td>
<td>✂</td>
<td>✓</td>
<td>✂</td>
</tr>
<tr>
<td>VCS</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>WAV</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>XLS/XLSX</td>
<td>✓</td>
<td>✓</td>
<td>?</td>
</tr>
<tr>
<td>XML</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
## Were File Formats Compliant?

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Email body</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>CSV</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>DOC/DOCX</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>EML</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>HTML</td>
<td>?</td>
<td>✓</td>
<td>?</td>
</tr>
<tr>
<td>ICS</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>JPEG</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>JSON</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>KMZ</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>MBOX</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>MP4</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>PDF</td>
<td>?</td>
<td>✓</td>
<td>?</td>
</tr>
<tr>
<td>PNG</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>RTF</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>TEX</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>TXT</td>
<td>?</td>
<td>✓</td>
<td>?</td>
</tr>
<tr>
<td>VCS</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>WAV</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>XLS/XLSX</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>XML</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Most formats either comply (✔️) or do not comply (✘) with the ICO definitions but some are ambiguous (❓).
Some data controller categories shown have at least one file format, with over 0.4 distribution represented in the responses.
Some data controller categories shown have at least one file format, with over 0.4 distribution represented in the responses.

File Formats within Data Controller Categories

Common file formats suggest the potential for establishing achievable standards on suitable data types for RtDP compliance.
OTHER PROBLEMS

Additional personal data required for identifying data subject: Photographic national ID, proof of address, recent bank transactions… – new privacy risks?
OTHER PROBLEMS

Confusion between different rights: Some data controllers thought we were making Article 15 (access) or Article 17 (erasure) requests

Hello Janis, thank you for your message. It appears that the previous e-mail was sent to you incorrectly and that your request is still being processed. I see that your account is still free to log in - thank you for your patience and understanding.

Sincerely,

Udacity Legal Team

On Thu, Aug 10, 2018 at 6:14 AM, Janis Wong wrote:
Dear Udacity Legal Team,

I never asked for my account to be deleted.

Kind Regards,

Janis

On Wed, 9 Aug 2018, 21:27 Legal Team, <privacy@udacity.com> wrote:
Dear Janis,

We have reviewed and verified your request for account deletion. This email confirms that the deletion has been completed. Please understand that this deletion is not reversible and may result in Udacity being unable to retrieve information about your account, enrollment, and records of completion. Please also keep in mind that all removals of such information are subject to requirements to maintain certain data in our archives for legal or legitimate business purposes.

Sincerely,

Udacity Legal Team
Security approaches varied: Data sent by e-mail (some of which were lost), passwords by post, CD-ROMs by post, download portals...
Data breach! One controller sent back data belonging to other requesters
OTHER PROBLEMS

Machine-readable? One controller claimed a machine-readable response was not possible; two sent paper scans as allegedly machine-readable

Hi Janis

Thank you for your email expressing you desire to exercise your 'Right to data portability' under article 20 of the General Data Protection Regulation (GDPR) 2016/679. I am unsure as to what data you think we may be holding at the Students Association that would be portable. I have check all our systems and records and can only see you connected to 1 transaction, which was for the purchase of a ticket to an event. The system that I can see it in is a read or refund only system therefore I am unable to give you the information within it in a Machine readable format.

If you require further information, please do not hesitate to contact me.

Kind Regards
Jillian

Lloyds DSAR – Scan Required

DSAR Reference:
WHAT NEXT?

- Looking at content of responses to determine compliance
- We need better definitions of structured, commonly used, and machine-readable
- Looking at the right to transmit (perhaps too early given lack of preparedness for even the easier bits of Art 20)
- Is interoperability feasible?
- How can technology help?
CONCLUSIONS

- GDPR makes it easier to conduct empirical data protection studies
- Data controllers vary in their approach to RtDP: file formats, security, interpretation of the requirements
- Better guidance needed
- More empirical work is needed