



**A mother cares for her sick son at Mahad camp for internally displaced persons in Juba, South Sudan.** Mahad camp is home to around 2,500 South Sudanese, mostly from Jonglei state, who arrived in February and March 2014. The conflict in South Sudan, which erupted in December 2013, has displaced more than 1.5 million individuals within the country.



# 6 From Data Protection to Statistics

**T**he demand and need for information about refugees and others of concern have grown over the years, particularly given the increase in the numbers of emergencies and protracted refugee situations, and the increasing use of the internet and electronic media. UNHCR and other humanitarian organizations have invested significant efforts in the collection and dissemination of data. This increasing demand for data, particularly at the individual level, has resulted in the need for

comprehensive and coherent data protection policies across the humanitarian community.

This chapter summarizes UNHCR's newly adopted *Policy on the Protection of Personal Data of Persons of Concern to UNHCR*<sup>98</sup> ('Data Protection Policy') and sets out its relevance to UNHCR's statistical and data analysis work. The relationship between data protection and statistics is not a new phenomenon and reference is also made to other relevant laws, resolutions, codes, and research publications.

## UNHCR's data protection policy

The protection of refugees' personal data or personally identifiable information is not new for UNHCR. As part of the organization's protection mandate, UNHCR many years ago integrated the principle of confidentiality of this information into its work and propagated it among States. However, the organization's recently issued Data Protection Policy is the first comprehensive and publicly available policy document on this issue. It introduces a number of concepts and internationally recognized principles.

Among the key concepts are broad definitions of 'personal data' and of the processing of this

data. Other key concepts include the distinction between a data controller and a data processor, what constitutes a personal data breach and the related duty to notify relevant institutions about such breaches, and data protection impact assessments. The policy identifies eight principles: legitimate and fair processing, purpose specification, necessity and proportionality, accuracy, respect for the rights of the data subject, confidentiality, security, and accountability and supervision.

<sup>98</sup> See UNHCR, *Policy on the Protection of Personal Data of Persons of Concern to UNHCR* (henceforth: UNHCR Data Protection Policy), May 2015, <http://www.refworld.org/docid/55643c1d4.html>.

## Personal data

The key concept of the Data Protection Policy is the definition of personal data. UNHCR defines personal data as ‘any data related to an individual who can be identified from that data; from that data and other information; or by means reasonably likely to be used related to that data.’<sup>99</sup> In essence, this definition follows the standard definition used in a number of legal instruments, including Council of Europe Convention No. 108 and EU Directive 94/46, both of which read ‘any information relating to an identified or identifiable individual (‘data subject’).’<sup>100</sup> The UNHCR definition simply describes the term ‘identifiable’ as ‘who can be identified ... from that data and other information; or by means reasonably likely to be used related to that data.’

The concept of personal data also determines the scope of the Data Protection Policy and its principles. The policy states that it ‘applies to all personal data held by UNHCR in relation to persons of concern to UNHCR. The processing of other data, e.g. aggregated or anonymized, does not fall within the scope of this Policy [...]’<sup>101</sup> Statistics contain by definition aggregated or numerical data, i.e. quantitative data consisting of numbers.<sup>102</sup> The Data Protection Policy would therefore normally not apply to statistics.

How does the Data Protection Policy apply to UNHCR’s statistics? First, let’s look at the principle of confidentiality. The policy states that

‘[p]ersonal data is by definition classified as confidential. The confidentiality of personal data must be respected by UNHCR when processing personal data at all times. In order to ensure and respect confidentiality, personal data must be filed and stored in a way that is accessible only to authorized personnel and transferred only through the use of protected means of communication.’<sup>103</sup>

Second, let’s look at the concept of a personal data breach. This is defined as ‘[a] breach of data security leading to the accidental or unlawful/illegitimate destruction, loss, alteration, unauthorized disclosure of, or access to, personal data transferred, stored or otherwise processed.’<sup>104</sup> Published statistics that contain individually identifiable information would be a personal data breach, in violation of the confidentiality principle. Moreover, a number of individual rights could hardly be respected, such as the data subject’s

rights to correction, deletion, and objection to the processing of their personal data.<sup>105</sup>

The Data Protection Policy therefore requires that statistics are made accessible to third parties if measures are taken to ensure that the data subjects are not identifiable. This requirement corresponds to an almost identical proposal made by the Committee of Ministers of the Council of Europe in its recommendation concerning the protection of personal data collected and processed for statistical purposes.<sup>106</sup> Personal data collected and processed for exclusively statistical purposes should not, in any circumstance, disclose the data subject’s identity, the Council of Ministers stated.<sup>107</sup>

This requirement can also be found in multiple ethical codes for statisticians.<sup>108</sup> For instance, the International Statistical Institute’s ‘Declaration on Professional Ethics’ sets standards regarding the obligations of statisticians to protect the identity of data subjects in the strongest way possible. The declaration makes particular mention of cases in which the means of publishing and releasing the data would allow for the identification of the individual.<sup>109</sup> In addition, when it comes to the publication of microdata, the ‘Microdata

<sup>99</sup> *Ibid.*, para. 1.4, Terms and Definitions.

<sup>100</sup> See Article 2 (a) of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data of 28 January 1981, available at: <http://conventions.coe.int/Treaty/en/Treaties/Html/108.htm>; Article 2 (a) of Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, available at: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML>.

<sup>101</sup> UNHCR Data Protection Policy, para. 1.3.1, Scope

<sup>102</sup> See <http://www.thefreedictionary.com/statistics>.

<sup>103</sup> UNHCR Data Protection Policy, para. 4.1, Confidentiality of personal data.

<sup>104</sup> *Ibid.*, para. 1.4, Terms and definitions.

<sup>105</sup> *Ibid.*, para. 3.3, Correction and deletion and 3.4 Objection.

<sup>106</sup> See Council of Europe Committee of Ministers, Rec(97)18E 30 September 1997 concerning the protection of personal data collected and processed for statistical purposes, para. 14, available at: <https://wcd.coe.int/com.instranet.InstraServlet?command=com.instranet.CmdBlobGet&InstranetImage=20017248&SecMode=1&DocId=578856&Usage=2>.

<sup>107</sup> See Council of Europe Committee of Ministers, Explanatory Memorandum, Rec(97)18E 30 September 1997 concerning the protection of personal data collected and processed for statistical purposes, p. 11-15, available at: [http://www.coe.int/t/dghl/standardsetting/dataprotection/EM/EM\\_R%2897%2918\\_EN.pdf](http://www.coe.int/t/dghl/standardsetting/dataprotection/EM/EM_R%2897%2918_EN.pdf).

<sup>108</sup> See International Statistical Institute (ISI), Declaration on Professional Ethics, p. 7, available at: <http://armstat.am/file/doc/99479658.pdf>; UNECE, Fundamental Principles of Official Statistics, available at: <http://www.unece.org/stats/archive/docs.fp.e.html>; Organization for Economic Co-operation and Development (OECD), Quality Framework and Guidelines for OECD Statistical Activities, available at: <http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=std/qfs%282011%291&doclanguage=en>; European Union (Eurostat), European Statistics Code of Practice, available at: <http://ec.europa.eu/eurostat/documents/3859598/5921861/K5-32-11-955-EN.PDF>.

<sup>109</sup> International Statistical Institute (ISI), Declaration on Professional Ethics, p. 7, available at: <http://armstat.am/file/doc/99479658.pdf>.

Dissemination Best Practices’ of the Committee for the Coordination of Statistical Activities lay out the principles behind and technical possibilities for the protection of individual identity in statistics.<sup>110</sup>

UNHCR rarely collects personal data for the sole purpose of publically available statistics. While the agency has a mandate to collect and disseminate statistics, the relevant paragraph in its Statute explicitly refers to the collection of numbers of refugees from governments.<sup>111</sup> As stated in

the Data Protection Policy, UNHCR, in pursuing its international protection and solutions mandate, is often required to process the personal data of persons of concern – for instance, for registration, status determination, the provision of assistance, protection monitoring, or organizing voluntary repatriation or resettlement. Even though the main purpose of UNHCR’s collection of personal data is not statistics, statistics are still a legitimate and specific aim for the agency.

## The Anonymization Challenge

In general, the potential means by which an individual could be identified, along with the availability of those means and the likeliness they will be used for this purpose, have to be taken into account when producing statistics. It is also important to consider all such means that could be used not just by ordinary people but also those who may have reason to purposefully try to identify an individual.<sup>112</sup>

Why is it so important to avoid the disclosure of personal data in UNHCR’s statistics? The personal data of persons of concern to the agency are considered to be generally sensitive,<sup>113</sup> as refugees and others under UNHCR’s mandate are often at risk of being persecuted or otherwise ill-treated.<sup>114</sup> As such, disclosure of personal data may result in harm to the refugee, family members, or other relatives.

In this context, statisticians will be particularly interested in what the definition of personal data has to say on information relating to identifiable individuals, especially with regard to more-detailed statistics. For example, a statistic that lists the number of refugees by country of origin in any given country of asylum could, in the case of very low numbers, be used to identify individuals.

In the absence of an internationally recognized convention on how to treat small numbers of individuals in a statistical pool, such numbers may pose a risk of identification in official statistics. Yet this risk can be reduced significantly by setting a ‘small-number threshold’.<sup>115</sup> As a small number of cases relating to a specific country of origin and destination country pose an increased

risk for an individual to be identified, it may be useful, for instance, to systematically remove such small numbers from official statistics. A major challenge with such an approach is the fact that these numbers would not appear within the statistics, and thus may distort the statistical picture. For this reason, the small-number threshold may not always be useful.

Another way of approaching small numbers is to replace them with an asterisk. This is a common practice in UNHCR’s official statistics, whereby figures between one and four are replaced with an asterisk in both its Population Statistics Database<sup>116</sup> and its downloadable spreadsheet files. Despite the agency’s commitment to and support for ‘open data’, it is nevertheless important to constantly review the data already disclosed and to analyse whether removing small numbers within the statistics is mandatory.

Eurostat compiles data related to asylum applications, decisions, and the backlog of undecided cases, among other issues, and rounds all such data on its website to the closest five. This means

<sup>110</sup> Committee for the Coordination of Statistical Activities, Microdata Dissemination best practices, available at: <http://unstats.un.org/unsd/accsub/2014docs-23rd/SA-2014-4-Microdata.pdf>.

<sup>111</sup> Statute of the Office of UN High Commissioner for Refugees, para. 8 (f), available at: <http://www.refworld.org/docid/3ae6b3628.html>.

<sup>112</sup> See <https://ico.org.uk/media/for-organisations/documents/1554/determining-what-is-personal-data.pdf>, p. 9.

<sup>113</sup> See UNHCR’s Data Protection Policy, para. 1.2.1, Rationale.

<sup>114</sup> See the definition of a refugee in Article 1 of the 1951 Convention as well as Article 1 of the 1969 OAU Convention.

<sup>115</sup> See [http://www.hscic.gov.uk/media/13158/Small-Numbers-Procedure/pdf/Small-Numbers\\_Procedure.pdf](http://www.hscic.gov.uk/media/13158/Small-Numbers-Procedure/pdf/Small-Numbers_Procedure.pdf), p. 6; <http://www.doh.wa.gov/Portals/1/Documents/5500/SmallNumbers.pdf>; <http://www.cohid.dphe.state.co.us/smnumguidelines.html#foot2>.

<sup>116</sup> See <http://popstats.unhcr.org/>.

that all figures are treated equally, irrespective of their magnitude. As a consequence, very small asylum figures ‘disappear’ on the Eurostat website. However, this approach poses an important challenge in the calculation of totals based on aggregated data.

An opposite trend has been observed in the United Kingdom in recent years. As with Eurostat, the British Home Office used to round all of its asylum statistics to the closest five before publishing. It explained its rationale for this approach as follows:

‘[D]espite the care which is taken in collecting and collating all the information obtained, the figures are subject to the inaccuracies inherent in any large recording system and are not necessarily accurate to the last digit. The rounding of figures also serves to ensure the confidentiality of the original source data used and the individuals to whom it relates.’<sup>117</sup>

However, this approach has gradually been revised to its current status, whereby nearly all of the Home Office’s aggregated asylum data is accessible on its website without any major restriction.

It is important to stress that the main purpose of statistics is not to single out an individual from data, but rather to show a particular trend or describe a situation. Situations in which there is even a slight

hypothetical possibility that an individual could be identified by reconstructing the statistics’ data should be avoided.<sup>118</sup> In the age of ‘big data’, it is often quite difficult to determine whether a possibility is merely hypothetical. Even if no means are reasonably likely to be used to identify an individual, it is still important to review the situation at a later stage in order to re-assess those means. After all, these can change over time given technological developments, newly published information, and changed legal constraints.<sup>119</sup>

UNHCR’s statistics do not contain identifiers that would make it possible to immediately link information to a specific human being by looking solely at the data. However, if some information provided by the agency’s statistics were to be combined with other information, a direct link could be made to a specific individual. In particular, this is the case with information about the location of refugees, their movement to another country, and specific characteristics (e.g. age, sex) when combined with statistics showing a low number of individuals. Removing all directly identifying elements from the statistics is not enough to prevent possible identification. Depending on the context and the purpose of the statistics, additional measures may be needed.<sup>120</sup>

## Conclusion

UNHCR’s Data Protection Policy is an important step under its mandate to ensure the protection of refugees, including refugees’ individual identity. When it comes to the production and publication of statistics that include data on refugees, special precautions need to be taken by statisticians in order to reduce the risk of identification of any individual. Because statistics can include information of varying sensitivity, and because these are structured and published or otherwise released in various ways, it is important to develop flexible dissemination policies to ensure the effectiveness of data protection policies. In view of the sensitivity of UNHCR’s data, this is particularly important for dealing with the agency’s statistics. ■

<sup>117</sup> See *Control of Immigration: Statistics United Kingdom 2006*, p. 111, available at [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/228967/7197.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/228967/7197.pdf)

<sup>118</sup> See Article 29 Data Protection Working Party, Opinion 4/2007 on the concept of personal data, [http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2007/wp136\\_en.pdf](http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2007/wp136_en.pdf), p. 15.

<sup>119</sup> Amendments of the European Parliament to the Draft General Data Protection Regulation, 21 October 2013, Recital 23: ‘To ascertain whether means are reasonably likely to be used to identify the individual, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration both available technology at the time of the processing and technological development.’

<sup>120</sup> See [http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216\\_en.pdf](http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf), p. 9.