



**STOCKHOLM INTERNATIONAL
PEACE RESEARCH INSTITUTE**

SIPRI data policy: principles and guidelines

I. General

The collection and analysis of data are an essential part of SIPRI's activities. The primary purpose of the SIPRI data is to support SIPRI's activities to monitor and analyse developments in armaments, arms control, disarmament, conflicts and peace operations with the overall aim to contribute to international peace and security. Fair use of the SIPRI data is not only permitted but also encouraged. However, for commercial use, SIPRI should recuperate some of the costs to SIPRI of generating the data.

These policy guidelines have generally been designed for digital, computer-based data sets but also apply to archive paper-based data sets.

II. Data ownership and governance

SIPRI data sets belong to SIPRI, not to the SIPRI employees or contract employees who have collected them during their SIPRI employment. Ownership of data sets implies the right to exploit them and the right to destroy them.

SIPRI researchers will often hold data owned by SIPRI. There are consequent obligations on the holders to look after the data responsibly and to ensure that assessing, managing, using, improving, monitoring, maintaining and protecting the data are regulated.

Researchers will often exchange data sets in a spirit of scientific cooperation, and it is a part of the normal process of research that they should do so. However, data sets should not be distributed to third parties without any restrictions on their subsequent re-use or redistribution. At times, therefore, data transfers need to be controlled by formal licence agreements.

Data sets can easily be destroyed unintentionally, or effectively lost, by failure to take adequate and continuing precautions to safeguard them. The following rules for data protection should therefore be observed:

1. It is mandatory to make regular backups of the data sets in two copies: one for internal storage and one for external storage.
2. Backups must be made in a system-independent form (for example SQL dumps) in order to be able to recreate them on other platforms if necessary.
3. A thorough documentation of the data set must be maintained and regularly updated. The documentation shall include a codebook as well as information on sources and methods.
4. Paper records and related documents must where possible be converted into electronic documents and integrated in the above-mentioned backup process.

III. Data licences

SIPRI does not sell data or the intellectual property rights to its data.

SIPRI does not make available entire data sets to other parties.

However, SIPRI will grant data licences and will accept unlicensed fair use of its data.

Fair usage

The fair use of SIPRI data is free of charge. Fair use exists:

- when data are used for research, education or reporting purposes but not exploited commercially;
- when data are used for purposes such as criticism, comment, news reporting, teaching or scholarship, in which the use is for non-commercial purposes;

- when the reproduction of data constitutes less than 10 per cent of the total data set.

Definitions of '10 per cent of the total data set' are to be made for each data set and published on project websites.

Non-commercial usage

It is SIPRI policy to enable access to data at no more than the direct costs involved. Depending on circumstances, this will be achieved by providing data either: wholly free of charge; for a nominal handling charge; or at an appropriately discounted rate. Such provision of data for non-commercial use must inevitably be based on mutual trust. SIPRI must be satisfied that data supplied free or at reduced cost for research will not be re-used for commercial purposes in breach of the licensing arrangements.

Partners in data acquisition will not be charged for unprocessed data, but where SIPRI has analysed, interpreted or generally added value to the data a charge for the resultant value-added information may be made.

SIPRI's data may be used as a potentially tradable asset for reciprocal exchange agreements between other research organizations.

The relevant SIPRI programme leader will normally receive the request for permission to use the data and make a recommendation for a decision to SIPRI's management, who will then undertake negotiations on data provision or data exchange (see also section IV).

Commercial usage

Commercial usage is the usage of SIPRI data to generate revenue.

Where a licensee intends to levy a charge, SIPRI will generally sell the right to use data (via a data licence) and provide access to the data so that the right can be exercised. However, no more than 50 per cent of an entire data set can be licensed to any one licensee.

It is SIPRI data policy to charge for the provision of data at a rate that is dependent on the use to which the data will be put and to specify formally any consequent restrictions on the use of the data in formal licensing agreements. Granting access to SIPRI data consists of a limited access to the SIPRI databases only, and it is a one-time delivery of a data set.

Costs for commercial usage

Unmodified data or pure subsets of SIPRI databases

Fees for unmodified data are calculated on a cost per unit basis.

The size of the updated data matrix for the most recent year is set in relation to the SIPRI production costs (e.g. salaries, overheads etc.) to arrive at costs for each cell of the data matrix (costs per unit). The number of data cells transferred to the licensee multiplied by the cost per unit (or a certain percentage of it) constitutes the fee (see the appendix).

Modified data

Modified data or specifically generated data have a mixed cost structure. One part consists of the above-mentioned costs for the transferred data cells of unmodified data and the other part of costs for data retrieval and programming.

Price categories

Different prices will be applied to the SIPRI data licensee depending on the nature of the data usage and their 'commercial' status (see the appendix). Prices of data licences will therefore differ accordingly. Each SIPRI database programme will provide an estimate of the cost per unit of their data. This unit cost will be used as a basis for calculating the price of the SIPRI data.

IV. Procedures and administration of data licensing

Handling of incoming data licensing requests

Incoming requests for all data licensing whether by the electronic request form, email, postage, telephone or fax are to be handled by the staff of the respective SIPRI research programme. They will assess the use and the user of the request.

The programme staff will then make a recommendation on the kind of licence that should be granted and what fees should be charged. This recommendation is then passed on to the administration department for further processing.

Processing of assessed data licensing requests

The administration department will then send the licence agreement form for a specific licence, along with an invoice, if relevant, to the data licensee. Upon return of the signed licence agreement and required payment, the staff of the research programme will make the data accessible for the licensee.

Granting access to licensed data

Data access is handled by the SIPRI research programme staff who grant access to the agreed subset of a database. A spreadsheet or similar electronic document will normally be sent to the licensee.

Appendix. Cost calculation and pricing guidelines

Calculating the costs per unit

The cost per data unit is equal to the sum of the (annual) costs for generating data (including salaries, overheads and IT costs) divided by the number of data units produced (per year).

$$c = \frac{F_1 + F_2 + \dots + F_n}{v_1 \times v_2 \times \dots \times v_m}$$

where

- c is the cost per data unit
- $F_1, F_2, \dots, F_n$ are the production factors (e.g. F_1 = salaries, F_1 = overheads, F_3 = IT costs, etc.), and
- $v_1, v_2, \dots, v_m$ are the determinants (variables) of the data matrix (e.g. v_1 = number of years, v_2 = number of countries, v_3 = number of GPD data, etc.).

Setting the price

The price should be set according to the principle of retrieving as much of the production cost as possible but not exceeding the market value of similar products. The data unit price would normally be set at 10–20 per cent of the data unit production costs.