

# **Open Access Statistics**

## **Data Provider for DSpace and WebDoc repositories**

Version 1.0

### **Documentation**



## 1 Introduction

---

The Open Access Statistics (OAS) Data Provider enables Open Access repositories to participate in the OAS infrastructure. It collects all necessary data by parsing the servers' log files, links it to the documents' metadata and provides it as Context Objects (XML) via OAI-PMH to be fetched by the central OAS Service Provider.

The Data Provider consists of two main modules:

- The Logfile Parser parses the server's log file, reads additional metadata from the repository and aggregates it into Context Objects which are cached in a local database.
- The Data Provider provides these Context Objects via a standardised OAI interface.

For detailed information about the project, please visit <http://www.dini.de/projekte/oa-statistik/english/>.

The OAS Data Provider is released under the MIT License. Some modules are licensed differently. For details, see *license.txt*.

PLEASE BE AWARE THAT THIS SOFTWARE IS NOT SUITABLE FOR PRODUCTIVE SYSTEMS AND SHOULD ONLY BE USED FOR TESTING PURPOSES. FURTHERMORE, TO FULFIL SEVERAL COUNTING STANDARDS LIKE COUNTER, IP ADDRESSES ARE LOGGED. EVEN THOUGH THEY ARE ENCRYPTED WITH A SALTED SHA256 HASH, AT LEAST IN GERMANY THERE IS AN ONGOING DEBATE IF THE LOGGING OF SUCH DATA WITHOUT THE USERS' CONSENT VIOLATES THE LAW.

OAS AND THE GÖTTINGEN STATE AND UNIVERSITY LIBRARY CANNOT BE MADE RESPONSIBLE IN ANY WAY FOR TECHNICAL OR LEGAL PROBLEMS THAT ARISE OUT OF THESE ISSUES.

## 2 System Requirements

---

The OAS Data Provider will be running on basically any "LAMP" system, i.e. a Linux distribution with the Apache web server and PHP under the following preconditions:

- PHP version 5.2.x or newer
  - Logfile Parser requires PHP being compiled with pcntl (note that the pcntl extension is not available for Windows) and CLI
  - Both modules require the DOM extension
- A database for storing Context Objects and metadata that is accessible via PDO (MySQL, SQLite or similar) with an account that has the rights to create tables
- Apache should be configured to log to file or, if unavoidable, to pipe

## 3 Installation

---

Just unpack the two main modules *logfile-parser* and *data-provider*.

The *logfile-parser* can be stored in any location, but should not be accessible from outside for privacy and security reasons. The *data-provider* should be stored in the document root and must be accessible from the outside, use *.htaccess* if necessary.

## 4 Configuration

---

For the DSpace and WebDoc repository servers, we provide the appropriate classes. Just edit the configuration files in the main directory *logfile-parser*, consisting of a common config file named *config.php* and a repository-specific one named *config-<name>.php*.

If you are running a different server, you may have to create these classes and config for yourself. The created Context Objects by *oasparser-webserver-standard.php* are lacking the ServiceType and any identifiers except an URI. So *logfile-parser/lib/oasparser-webserver-standard.php* has to be extended by a repository-specific class in *logfile-parser/lib/oasparser-webserver-<name>.php* which in turn requires its own library *logfile-parser/lib/identifiers/lib/lib-<name>.php* to collect data that can not be obtained from the server's logfile and of course a configuration file named *logfile-parser/config-<name>.php*. Include *lib-<name>.php* via "require" in your config file and set "parser" accordingly.

When ready, initialize the database by calling *log2ctx -O*. The *config.php* is used by default, use *log2ctx -c* for a custom config file. The parameter *-i* is used to set an identifier string. See *readme.txt* for details.

There are two unix shell scripts bundled with logfile-parser:

- For a complete example on how to harvest metadata and convert webserver logfiles to Context Objects, see *update-script.sh*.
- To get rid of old records, use *cleanup-script.sh*.

Both scripts should be called regularly by a cronjob.

---

## 5 Contact Information

---

If you have any problems or questions regarding the configuration and operation of the OAS Data Provider for DSpace and WebDoc repositories, please feel free to contact us any time. We will try to answer your questions within a few days.

- [Hans-Werner Hilse](#)
- [Tobias Schäfer](#)