NatHosting – Technical Handout for Publishers

NatHosting is archiving metadata and full texts from publishers in a Private LOCKSS Network (PLN) to ensure their sustainable availability.

A publisher either provides content via File Transfer Protocol (FTP) or allows the NatHosting PLN to harvest content from the publisher’s website.

The NatHosting Private LOCKSS Network

The NatHosting Private LOCKSS Network – in contrast to the Global LOCKSS Network – is operated by NatHosting partners exclusively and content will be available only to entitled users. It consists of several servers running the LOCKSS services – so called LOCKSS nodes. These nodes are operated by NatHosting project partners at different sites. FIZ Karlsruhe is in charge of the evaluation of the software, the documentation of the installation, operation and network infrastructure and for technical support. No fees will be charged to publishers.

Copies of the ingested content will be stored at several nodes of the NatHosting Private LOCKSS Network. These nodes communicate among each other to regularly check the comprehensiveness and the integrity of the data. Lost or corrupt data at one node will be replaced by valid data from other nodes.

Option 1: Supply via FTP

A publisher may send content to an FTP server operated by NatHosting or provide content on its own FTP server. Data packages containing full texts and metadata are transferred via File Transfer Protocol in this case. The structure of the metadata and their relation to the respective full texts has to be uniform for all publications within a distinct delivery in order to be processable by NatHosting.

Option 2: Websites harvested by NatHosting

In order to load content from the publisher’s website to the NatHosting PLN, content has to be accessible to NatHosting without password authentication. Usually, a publisher allows NatHosting nodes access after verifying their IP addresses.

In addition, a so-called permission statement has to be placed on the publisher’s website referring to the content to be harvested. NatHosting thus gets the information that the content may be harvested and stored in the PLN. If no permission statement is found, NatHosting will not harvest the corresponding content.

The permission statement says "LOCKSS system has permission to collect, preserve, and serve this Archival Unit." Only with this statement the publisher explicitly allows access to the relevant content. The statement can be posted in the source code of a webpage, invisible to a general user of the website. For further details, see https://lockss.github.io/web-harvest-guidelines.html#permissions-statement
**Metadata and Full Texts**

All full texts within a supplied data package or to be harvested from the website as well as the related metadata have to be unambiguously identifiable.

Full text publications will be stored in the delivered format. There will be no format conversion. The metadata for journals should contain e.g. the following fields:

- Publisher name
- Title of publication
- Volume
- Issue
- ISSN
- Publication date
- Journal URL
- Article title
- Author(s)
- DOI
- Article URL

For technical questions, please, contact Frank Schwichtenberg (Frank.Schwichtenberg@fiz-karlsruhe.de).