

# ReadSpeaker docReader and Storage of Personal Data

This document describes how the web application service ReadSpeaker docReader (docReader for short) handles personal data.

Details may change in future versions but will essentially remain the same.

## How docReader Works

docReader is launched from a link on the website or web platform where it is implemented. The customer must define which document docReader should read (the target document). Clicking the link will open the target document in the docReader user interface in a separate browser window or frame, depending on the implementation. docReader will:

1. Check if the document has previously been cached. If so, skip to step 4.
2. Download a copy of the target document.
3. Extract the content of the document.
4. Present the document content to the user in a web interface.

All communication to and from docReader must be done over an encrypted channel using HTTPS (TLS  $\geq$ 1.2).

Content designated for text to speech is processed in the ReadSpeaker environment, converting the content to speech. When the audio has been generated, a redirect is sent back to the user which triggers the web browser to make a request to a server where the generated audio resides.

All data that is processed during audio generation is stored on encrypted disks using modern ciphers that are known to be secure.

## Data Centers

ReadSpeaker has its own equipment at two data centers located in Stockholm/Sweden and Ashburn VA/USA. The data centers mainly serve their respective regions. Additional data centers may be set up in the future.

Customers choose which data center they want to use when implementing the service. The docReader service can be configured to automatically failover to another data center if the chosen one becomes unreachable, except where prohibited by local regulations.

We use several external CDN vendors to distribute static application files. See section [CDN Vendors](#) below.

## Collected, Stored, and Used Data

End user data such as user agent, OS type, device type, cookies, referrer data, IP address, and geolocation is stored only in temporary web server logs.

If personal information is present in a document and is available to be read, it can be displayed on the screen and sent over to our data centers to be turned into audio files. If there is no personal data in the document, no personal data will be processed by docReader.

Three kinds of logs are used for different purposes: web server logs, statistics logs, and an application log. All logs are stored using modern and secure encryption.

## Web Server Logs

These logs are necessary for troubleshooting and to prevent illegitimate usage. The log files are standard web server logs. Access to these logs is restricted to a limited number of people with specific roles related to maintenance. Web server log records older than 30 days are deleted permanently.

The log information is restricted to technical staff and only reachable through encrypted channels using MFA. ReadSpeaker does not sell or pass on any log information.

## Statistics Logs

The data for statistics logs is anonymized directly. It contains customer ID, timestamps, the requested URL, the length of the text content, and which voice was used. This data is used for billing.

Statistics are saved per activation. An activation is when a user clicks the Listen button to get content read aloud.

The statistics log contains the URL the button was activated from. Unless the URLs are unique for each user, no user info is saved. If that should be the case, the implementer of the docReader service can override what should be stored so that the URL does not get stored.

## Application Log

Error and warning notices are stored in the application log, to notify ReadSpeaker of application issues and facilitate troubleshooting. Information stored in the application log contains timestamp, the URL of the target document, referrer, and error/warning details. Care is taken to exclude or anonymize any personal data.

## External Party/Subcontractor

### Translation Feature

The translation feature uses APIs from Google Cloud or DeepL to translate content. Which provider is used is configurable per customer.

All translation requests, regardless of provider, go through ReadSpeaker's servers. Only the content to be translated is transferred to the third-party API, never IP addresses or other personal information belonging to the end user.

If a customer does not want the translation feature to be available in docReader, it can be turned off in the customer's configuration.

More information on the handling of data by Google can be found on the Google Cloud Translation API help pages <https://cloud.google.com/translate/data-usage>.

More information on the handling of data by DeepL can be found on <https://www.deepl.com/en/privacy>.

### Dictionary Feature

The dictionary feature uses APIs from Oxford Dictionaries, Wiktionary, or dictionaries running locally in the ReadSpeaker environment to look up definitions of words. Which provider is used depends on the language.

All dictionary requests, regardless of provider, go through ReadSpeaker's servers. Only the single word to be looked up is transferred to the third-party API, never IP addresses or other personal information belonging to the end user.

If a customer does not want the dictionary feature to be available in docReader, it can be turned off in the customer's configuration.

### OCR Feature

The OCR feature uses the ABBYY Cloud OCR SDK to extract text content from images. Pages from the target document will be uploaded to the OCR service on a page-by-page basis as they are requested by the user.

The OCR feature is not enabled by default, but must be purchased separately.

More information on the handling of data by ABBYY can be found on <https://www.abbyy.com/legal/cloud-terms-of-service/dpa-universal-data-processing/>

## CDN Vendors

For the distribution of static files such as JavaScript, CSS, media files, etc., we use external content delivery network (CDN) providers.

The distributed files are required for the functionality of the service. When an end user's browser fetches files from the CDN, the IP, the user agent, and possibly the referer are unavoidably sent to the CDN. However, the customer can always choose a CDN where traffic remains within the specified region. Nodes within regions are selected to meet the requirements of regional privacy regulations. ReadSpeaker makes sure to provide CDNs for both global storage and regional storage, depending on what control the customer wishes to have over where the information shall reside.

The distributed files do not contain any content from the customer's website. Also, they cannot be used in any way to identify or gather any personal information and are therefore safe to distribute publicly over the Internet.

## Caching

docReader holds a copy of converted documents, or the original documents depending on the format, on disk for up to 30 days to extract parts of the document on demand and to speed up subsequent requests for the same document.

Content and metadata extracted from documents is stored for up to 30 days to enhance performance.

For better performance, docReader may also cache generated audio. By default, this is only temporarily stored on disk. Cache clearing is done individually for each file. A file's caching time is extended as long as it continues to receive requests.

While the exact caching duration may be adjusted for capacity management purposes, cached content is never stored for more than 30 days.

## Distribution to Third Parties

There is no distribution of collected data, apart from to the subcontractors mentioned above.

## Cookies

ReadSpeaker docReader stores one cookie in the user's browser:

1. dcrsettings - This is used for storing user settings so that they do not need to be reapplied on every page load. This cookie has a default expiration of 30 days.

Data purposes: docReader saves the user's settings in a cookie, so they are remembered and applied when the user navigates between pages.

Data collected: No data is collected. Information about reading speed, highlighting colors, etc. are saved in the cookies. No IP address is stored.

Data recipients: No data from the cookie is sent anywhere, it's only stored in the browser.

## Local Storage

docReader's highlighter feature stores information in the browser's local storage. Separate entries will be made for each document where the highlighter feature is used. Storing data has two purposes:

1. Storing settings related to the highlighter tool, such as the collection scope, sort order, and whether highlight borders should be displayed.

2. Storing the text ranges that should be highlighted and the colors associated with them. The text content of each highlighted range is also stored.

The highlight information in local storage is saved until the user clears all highlights from a document, or until the local storage is cleared from the browser.

Highlighter settings are stored until they are cleared from the browser, as they contain no personal information.

## GDPR

Since May 2018, ReadSpeaker has been committed to ensuring full compliance with the EU General Data Protection Regulation (GDPR).

## Data Privacy Matters

For inquiries regarding privacy matters, please contact our data protection team at [gdpr@readspeaker.com](mailto:gdpr@readspeaker.com).

## Further Questions

If you have further questions, please contact our support team via email, [support@readspeaker.com](mailto:support@readspeaker.com).