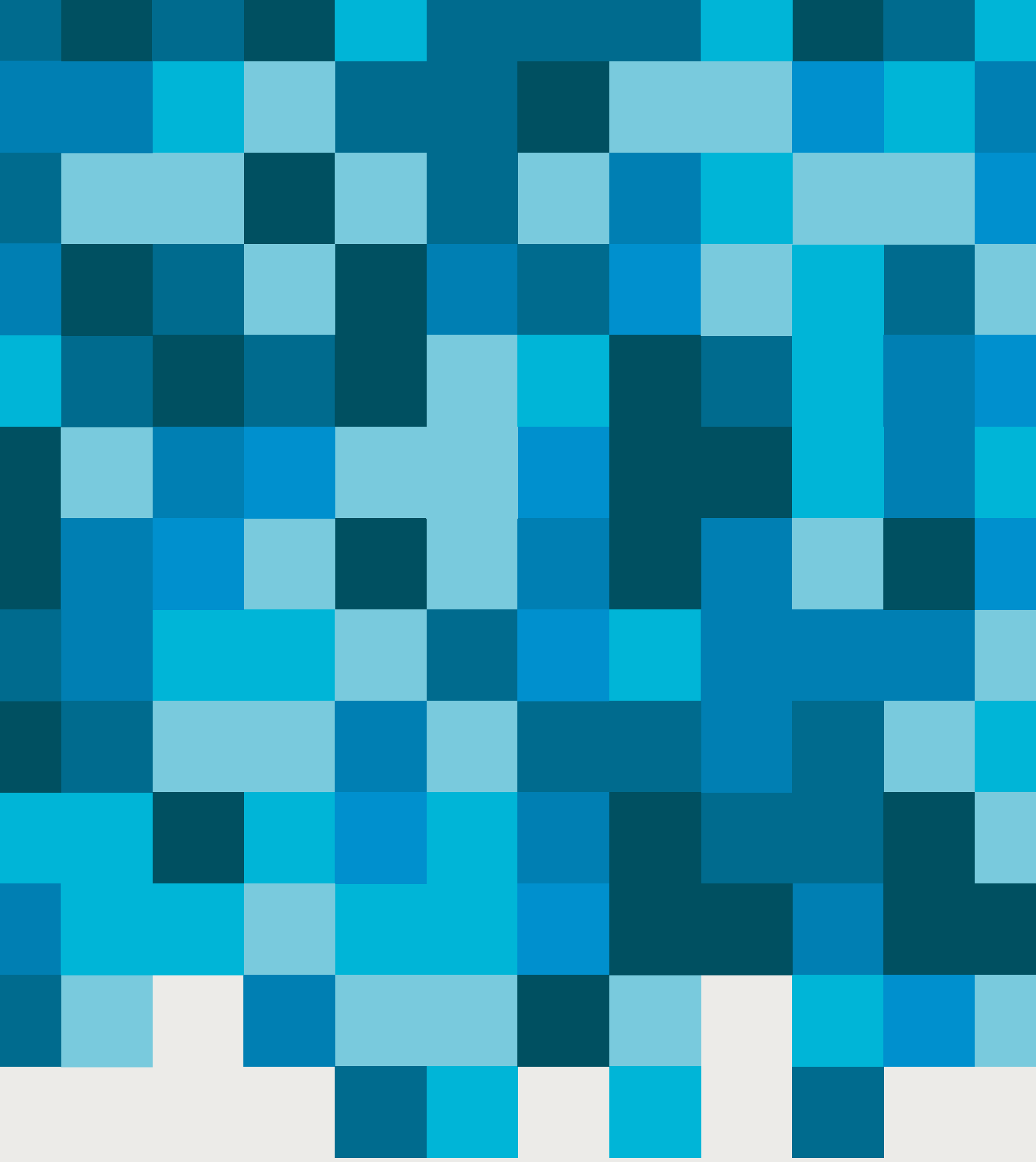
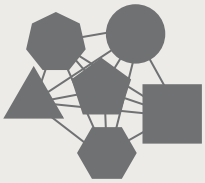


docWorks
Professional conversion software

CCS

docWorks
One workflow, no barriers.

docWorks is the first software that converts scanned pages to searchable and metadata enriched digital objects in one seamless workflow. From the import of the scans to the export as standardized formats, libraries and archives stay in full control of their files and do not have to deal with incompatibilities of modules or lost data shipments.


Other conversion solutions

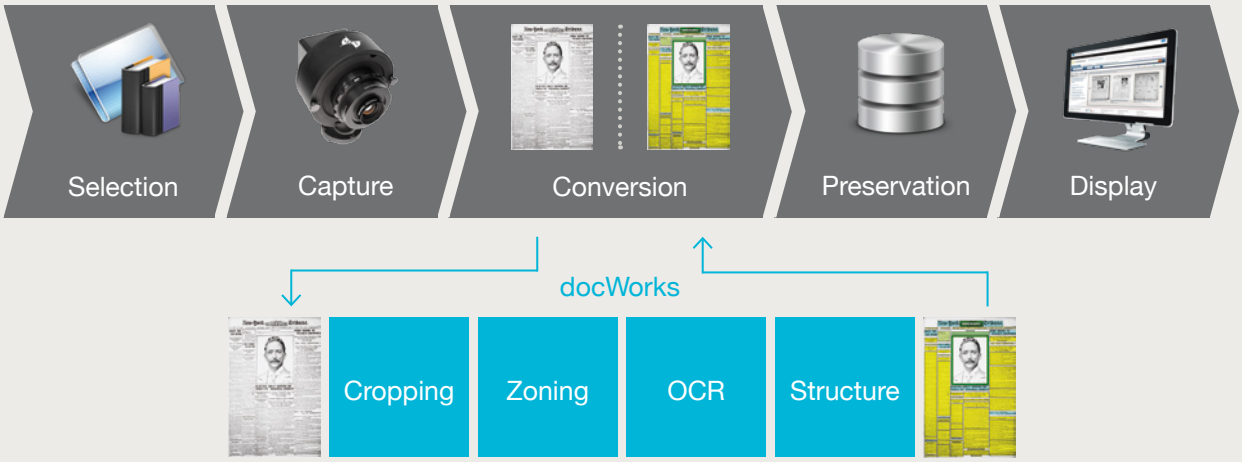

docWorks workflow

docWorks is used by the leading libraries and archives in the world to create digital collections. Clients include the Library of Congress, the British Library, the National Library of Australia, Harvard and Princeton University, Leica Camera Archives as well as many other national libraries, universities, content providers and cultural heritage institutions.

Digital collections
Enrich life by digitizing.

The reasons that libraries and archives digitize their holdings vary, but two motivations recur: The wish to grant an easy and worldwide access to knowledge and the digital preservation of this knowledge for future generations.

When creating a digital collection the workflow usually follows the below pictured sequence. It starts with the selection of the to-be-digitized items and ends with the storage and possible display of the digital objects. It is important to note that high quality conversion starts with preservation grade image capture. The better the input, the more successful the OCR.



Conversion
Giving sense to digital collections.

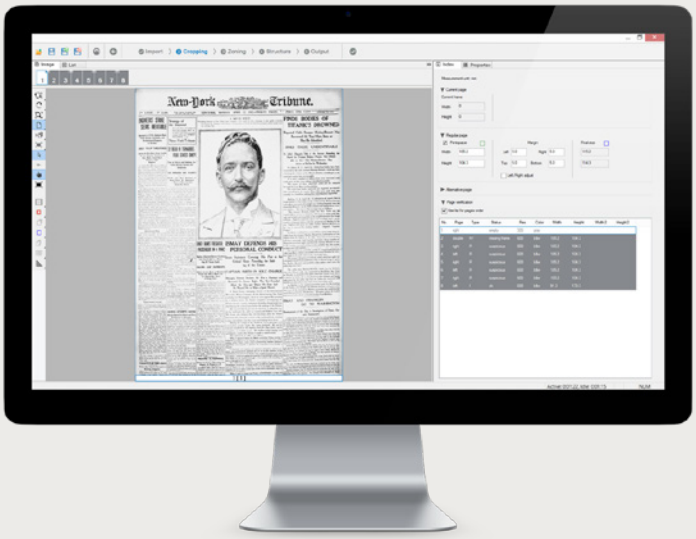
Conversion means the extraction of the content that is “sealed” inside the scanned pages. A scanned page is merely a photograph of the original print page. All the extensible content of the original print version such as text, illustrations and captions is not captured. To retrieve this information and make the publication fully available for search and database integration, libraries and archives use docWorks.
docWorks runs four steps to convert the “flat” scan image in a content enriched digital object.

1 *Cropping* is used for possible image corrections. If necessary pages can be straightened or cut to a consistent size.

2 Cropping is followed by *Zoning* where the content of each page (articles, illustrations, captions) is detected and identified. Breaking down the page into various content blocks is important especially for newspapers with their mixed layouts.

3 When running *OCR* over the page, done in docWorks’ next conversion step, the extracted text can be matched to each article. If there were no previous zoning, the whole text of a newspaper page would be just one single running text of intermingled articles.

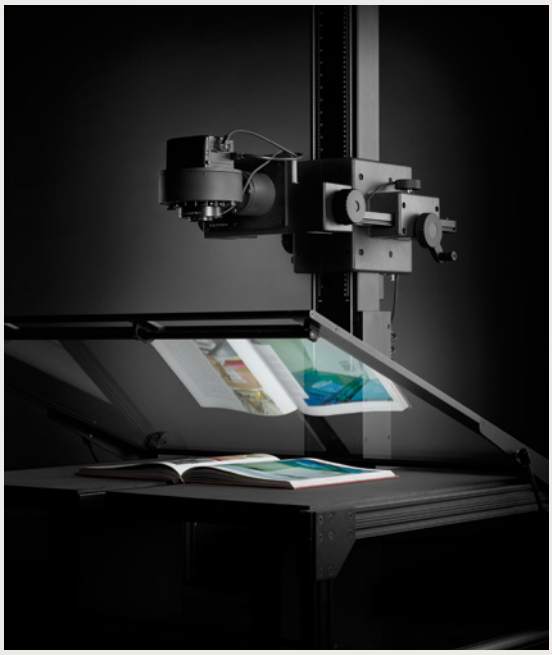
4 The final step is *Structure*. Here all pages are allotted a page type within the overall structure of the publication to specify their character as index, chapter or appendix.



docWorks interface
Order to complexity.

The docWorks interface was designed to be intuitive and easy to learn even for beginners. At the center of the interface is a step-by-step workflow that consists of the previously described conversion steps. Each step leads to the next one guiding the user gently through the whole conversion process. Brief statements for each step give assistance and background information if needed. The user interface has a modern, clean and pared-down design with carefully applied coloring to highlight specific function and ease the navigation.

From docWorks' Control Center you can constantly monitor your projects and configure the settings of your docWorks setup.



DT RGC180 Capture Cradle

DT BC100 Book Capture System

docWorks formats

Input	TIFF (uncompressed, group 3/4, packbits, lzw) JPG, JP2, BMP, GIF and PDF Color (24bit), Gray (8bit), Black/White (1bit)
Output	METS XML and ALTO XML Images (JPG, JP2, TIFF or others) PDF, PDF/A 1a and 1b Full text XML and RTF files, ePub
Metadata	MIX for technical metadata on images METS:structural map type="physical" METS:structural map type="logical" MARC, MODS and DC as describing metadata (possible to be imported by Z39.50 interface)
	docWorks runs on Windows 7, 8 and XP.

docWorks summary

- Conversion software for libraries and archives
- 35 years and 150 million processed pages of experience
- Used by the world's elite libraries and archives
- Barrier-free conversion process
- Clean and discreet interface
- Intuitive and assisting workflow
- Control and monitoring center
- Standardized output

docWorks editions
The right fit.

docWorks comes in four different editions. Each edition is directed towards a specific field of application, from the affordable Starter edition for small digitization projects to the flagship Enterprise edition for very large and ambitious collections. If projects are upscaled docWorks can easily be adapted by upgrading to a higher edition.

docWorks Starter edition
docWorks Starter is the perfect fit for small digitization projects. It runs on a single workstation and is easy to use even for inexperienced operators. Organizations which are scanning items or already have digital data available can use docWorks Starter to create METS/ALTO output from their PDF and ePub files.
Comprehensive video tutorials are available from CCS via its support web platform.

docWorks Basic edition
docWorks Basic is the next level of docWorks and is used for mid-size digitization projects. docWorks Basic is a client/server ready solution running docWorks on 2 computers in parallel. This way the workload can be split across two operators which maximizes the speed of performance. All automated processes are run by a server in the background.

If you want to add more work stations to your Basic edition you can order extra licences anytime - or of course upgrade to a higher docWorks edition.

docWorks Pro edition
docWorks Pro is used for professional digitization projects with a large volume of pages per year. It is typically installed if there is a multi-year digitization plan or a big project. The setup consists of 4 workstation licences plus the module *ScanClient*, which can be used on scanner computers to easily transfer images and import documents into the docWorks process. Two servers ensure smooth background processing.

In addition to docWorks's standard OCR software *Tesseract* the Pro edition also comes with *Finereader* OCR by ABBYY.

docWorks Enterprise edition
The docWorks Enterprise edition is the flagship of docWorks. It is used by the world's elite of libraries and archives to convert very large and valuable holdings into professional digital collections. docWorks Enterprise includes licences for 20 work stations, 8 servers and 2 ScanClients. CCS will consult your IT department about hardware and helps with the software integration to your library system.

---

docWorks User Group
Discuss with the like-minded.

All our docWorks clients are also part of the docWorks User Group. The members of the user group meet regularly with docWorks staff in online or face-to-face conferences. This way all the docWorks users get to know each other and can exchange experiences as well as receive best practice information first-hand from CCS staff.

Contact

U.S. distributor of docWorks

Digital Transitions Division of Cultural Heritage
35 West 35th Street, 4th Floor
New York, NY 10001
United States
T +1 212 529 6825
F +1 212 504 2713
E info@dtdch.com
W www.dtdch.com

Manufacturer of docWorks

Content Conversion Specialists GmbH
Weidestr. 134
22083 Hamburg
Germany
T +49 40 227 130-0
F +49 40 227 130-11
E info@content-conversion.com
W www.content-conversion.com