

Onboarding

BitCurator projects are maintained by members of the community. Interested in being added as a member? Read about areas you can contribute to and the technologies used below! Contact with our Software Development Committee

The code and development documentation for all BitCurator projects are maintained in Github Repositories (<https://github.com/bitcurator>), where both non-coding and development work is maintained and completed.

More information on using Git and Github:

- [Git Handbook · GitHub Guides](#)

Non-coding

There are a variety of tasks essential to the development and maintenance of BCC projects that don't require prior experience with coding.

Testing software, documenting bugs:

- [Top 7 Major Goals of A Good Software Tester](#)
- [How to Write A Good Bug Report? Tips and Tricks](#)
- [Bug report writing guidelines - Mozilla | MDN](#)

Check and Confirm bugs that been submitted as issues

- <https://github.com/BitCurator/bitcurator-distro/issues>
- <https://github.com/BitCurator/bitcurator-distro-installer/issues>
- <https://github.com/BitCurator/bitcurator-distro-tools/issues>
- <https://github.com/BitCurator/bitcurator-distro-metadata/issues>
- <https://github.com/BitCurator/bitcurator-access-webtools/issues>
- <https://github.com/BitCurator/bitcurator-access-redaction/issues>
- <https://github.com/BitCurator/bitcurator-redact-pdf/issues>
- <https://github.com/BitCurator/bitcurator-nlp-gentm/issues>
- <https://github.com/BitCurator/bitcurator-nlp-entspan/issues>

Documentation - Take part in keeping project user and development documentation up-to-date.

- <https://confluence.educopia.org/display/BC/Development>
- <https://github.com/BitCurator/bitcurator-distro/wiki>
- <https://github.com/BitCurator/bitcurator-access/wiki>
- <https://github.com/BitCurator/bitcurator-distro/blob/master/README.md>
- <https://github.com/BitCurator/bitcurator-distro-salt/blob/master/README.md>
- <https://github.com/BitCurator/bitcurator-distro-tools/blob/master/README.md>
- <https://github.com/BitCurator/bitcurator-distro-metadata/blob/master/README.md>
- <https://github.com/BitCurator/bitcurator-access-webtools/blob/master/README.md>
- <https://github.com/BitCurator/bitcurator-access-redaction/blob/master/README.md>
- <https://github.com/BitCurator/bitcurator-redact-pdf/blob/master/README.md>
- <https://github.com/BitCurator/bitcurator-nlp-gentm/blob/master/README.md>
- <https://github.com/BitCurator/bitcurator-nlp-entspan/blob/master/README.md>

For developers

The BitCurator family of research projects include three main projects: BitCurator, BitCurator Access, and BitCurator NLP. The repositories for each are linked below along with links to documentation for working with the technology stacks used in each project.

BitCurator Environment

The BitCurator environment is a customized variant of Ubuntu, configured using Salt Stack along with several repositories of custom tools including: bitcurator-adduser and bitcurator-metadata

Current repositories:

- <https://github.com/BitCurator/bitcurator-distro>
- <https://github.com/BitCurator/bitcurator-distro-installer>

- <https://github.com/BitCurator/bitcurator-distro-salt>
- <https://github.com/BitCurator/bitcurator-distro-tools>

Optional functionality and scripts:

- <https://github.com/BitCurator/bitcurator-distro-adduser> <https://github.com/BitCurator/bitcurator-distro-metadata>

Technology Stack:

- Ubuntu - Base operating system, user interface, tools
 - help.ubuntu.com
- Salt Stack -
 - [SaltStack Documentation](#)
- Shell Scripting -
 - [BASH Programming - Introduction HOW-TO](#)
 - [Bash Scripting Tutorial for Beginners](#)
- Python -
 - [Our Documentation - python.org](#)
- Related components -
 - [reingart/pyfpdf: Simple PDF generation for Python \(FPDF PHP port\)](#)
- XSLT -
 - [Introduction to XSLT](#)

BitCurator Access Webtools

The BitCurator Access Webtools project is comprised of a single repository. The README provides instructions for both end users and developers to clone and build from source.

Current repository:

- <https://github.com/BitCurator/bitcurator-access-webtools>

Technology Stack:

- The Sleuth Kit -
 - <http://www.sleuthkit.org/sleuthkit/>
- PyTSK -
 - <https://github.com/py4n6/pytsk>
- Flask -
 - <https://flask.pocoo.org/>
- Postgres -
 - <https://www.postgresql.org/docs/current/>

BitCurator Access Redaction Tools

The BitCurator Access Redaction tools project is comprised of two repositories. The READMEs provide instructions for both end users and developers to clone and build from source.

Current repositories:

- <https://github.com/BitCurator/bitcurator-access-redaction>
- <https://github.com/BitCurator/bitcurator-redact-pdf>

Technology Stack:

- The Sleuth Kit -
 - <http://www.sleuthkit.org/sleuthkit/>
- Liblightgrep -
 - <http://strozfriedberg.github.io/liblightgrep/>

BitCurator NLP Tools

The BitCurator NLP project includes several repositories. The topic model generation environment (bitcurator-nlp-gentm) enables automatic extraction of text from heterogeneous document collections within disk images to generate user-browsable topic models within a web browser. The disk browsing environment (bitcurator-access-webtools) provides full-text browsing of documents contained within disk images, along with (in progress) analysis of entities identified within those documents. Various command-line tools are provided in another repository (bitcurator-nlp-entspan).

Current repositories:

- <https://github.com/BitCurator/bitcurator-nlp-gentm>
- <https://github.com/BitCurator/bitcurator-access-webtools>
- <https://github.com/BitCurator/bitcurator-nlp-entspan>

Technology Stack:

- Sleuth Kit - to parse file systems in disk images,
 - <https://github.com/sleuthkit/sleuthkit>
- textract - to extract text from common file formats

- <https://textract.readthedocs.io/en/stable/>
- gensim - to generate topic models
 - <https://radimrehurek.com/gensim/>
- pyLDAvis -
 - <https://github.com/bmabey/pyLDAvis>)