# The Open Access Guerilla Cookbook

Dedicated to Aaron Swarz (1986-2013)

## Introduction

In 2008, a short and passionate appeal appeared online which called for all of us to, "take information, wherever it is stored, make our copies and share them with the world. We need to take stuff that's out of copyright and add it to the archive. We need to buy secret databases and put them on the Web. We need to download scientific journals and upload them to file sharing networks. We need to fight for Guerilla Open Access."

In January, 2013, its author Aaron Swarz took his own life. Among his many achievements included an important initiative to liberate the PACER court records database and leadership roles in several movements that support free culture. In the last years of his life, he faced serious criminal charges for following his own manifesto and launching one of the largest content liberation efforts to date: the downloading of millions of articles from the JSTOR database.

If the threat of decades of prison time for the JSTOR raid was designed to strike fear into OA guerillas everywhere, the tragic death of this selfless supporter of the movement should be met with a renewed commitment to our ideals.  It also offers us an opportunity to up our game.

This document aims to take the manifesto to the next step. This first version is merely an opening call, and it is in no way complete. Hopefully, this cookbook will grow to include many recipes and instructional essays for use by the open access guerilla. Improve on it, share it, and use it.

## Principles

The guerilla open access movement is founded on the basic tenant that the efforts to promote open access by operating exclusively within existing copyright regimes and attempts to reform these copyright regimes through legal reform do not go far enough to protect and expand the realm of free culture. Working in parallel, but often in the face of criticism from those promoting legal means to achieve open access, our guerilla movement accepts the active violation of copyrights and contractual terms of use as justified. We promote the mass liberation of content from commercial as well as non-profit or governmental databases for the greater purpose of sharing knowledge and culture with everyone.

We are closely allied with movements promoting government and corporate transparency, but open access guerillas recognize that there is responsibility in sharing information. When relevant, precautions should be taken to defend the safety and privacy of individuals and communities in appropriate ways.

We are pirates, but accept a moral imperative to loot more than we need for our own purposes, and share widely everything we find. We categorically reject descriptions of our acts as simple theft. We do not deprive humanity of culture, we reproduce it. We do not rob owners of their property, but in many cases violate their temporary and exclusive monopoly to profit by it in an age when the reasonable limits first set upon this monopoly have been long forgotten. We violate crippling terms of use on content in an age when access to almost all information requires accepting contractual limitations that few ever read. We re-release materials that are already in the public domain but locked behind paywalls or targetted by copyfraud.

Looking forward, the strength of the guerilla open access movement depends on combining efforts both secret and open, both collaborative and individual. We must:

**Share skills and experiences as well as content**

Content liberation should not solely depend on a few individuals with highly technical knowledge. We must work harder to share our skills and our experiences widely. This should include efforts to better reach out to those who are new but eager to learn the more challenging technical side of our guerilla efforts. We should also work towards establishing standards in the quality of our content collections, security practices to protect our illegal efforts, and a code of ethics for operating with the content sources we raid.

**Recognize a diversity of roles and a diversity of approaches**

We must abandon the image of the lone hacker as the symbol of our movement and recognize that any successful guerilla movement depends on the work of people filling many different roles. Some of these are described below.

We are radical in our means, dedicated to our ideals, and will be reviled and ridiculed by many. Many with similar goals to our own reject us, but they should still be considered as allies. Our movement exists within an ecology of culture creation, curation, and consumption. We must respect everyone who plays a role in the interdependent whole, even as we oppose the legal regime under which they operate. There are many artists, writers, scholars, archivists, librarians, developers, and non-profit organizations who strongly oppose us. They argue that their livelihoods are threatened by our actions, while others secretely sympathize with us. Let us be mature enough to admit that some of our targets produced their collections of content with little funding, charge their access fees with no mind to profit themselves, and host servers with a barebones maintenance staff. When we liberate their content, or share with others, keep always in mind the work that was put into creating and publishing it. Show as much respect as is compatible with our goals and remember that we are nothing without them.

**Collaborate to reduce risk and maximize scale**

We now live in the world of crowdsourcing. The power of a lone hacker armed with a scraper and some knowledge of security is not to be underestimated, but it comes at great risk of discovery and sacrifice. We must explore ways to better combine our efforts. The RECAP effort for the PACER archive of court documents is one model of how this approach works within the legal realm, we can learn from it and others. We should develop the means to conceal systematic content liberation in the invisible mass of everyday consumption. Rather than grabbing whole archives and document databases at once, we should take them in smaller pieces, with care to preserve their metadata integrity, and plans in place to reassemble the whole when an operation is complete.

**Segregate open from secret action**

Aaron Swarz combined strong public advocacy with secret guerilla action. In his case, it was not key to his discovery, but it is likely to have impacted the severity of the charges brought against him. In all guerilla movements it is important to segregate open from secret action. It is unwise to be an open voice for radical illegal action and also its agent. If you begin engaging in dangerous OA guerilla action, temper your public voice and avoid drawing attention to yourself, especially with regard to the virtues of illegal content liberation.

**Protect the Public Domain First**

Almost none of us are against the principle of limited terms of copyright as a way to promote the long term expansion of the public domain. The passion that drives most us to these radical measures would not exist in a world with copyright terms of ten or fifteen years. At this writing, there

are very few signs that legal reforms will move us back to these limits, and on the contrary, in many places around the world the trend is in the opposite direction.

More threatening, however, is the assault upon what is already in the public domain. In acts of copyfraud, publishers and digital service providers claim rights on content they do not own. They claim that new rights are produced in their digitization and, increasingly, they are moving away from copyright to contractual terms of use to limit our freedom to use public domain materials obtained from their databases. The ever increasing proportion of our cultural heritage bound by these contractual restrictions will have direct consequences. The trend significantly drains support for initiatives to create fully open and free databases of materials that may already have been made available by more restrictive ventures, whether they are commercial or non-profit.

For these reasons, our movement's priority should be on the liberation of content in the public domain followed by those materials that, by any just limited term of copyright, should have in the public domain decades ago.

## Roles in the Movement

There are many ways to further the goals of the guerilla open access movement. Find one or several of the following roles that you feel comfortable performing.

**The Advocate**

The advocate promotes the cause of open access. Many OA and copyright reform advocates believe we rob their efforts of legitimacy by making it easier for content industries to smear them with our sins. Others believe, on principal, that OA must depend always on voluntary sharing. As one leader in the legal OA movement puts it, "There is no vigilante OA, no infringing, expropriating, or piratical OA." We disagree. We believe our efforts usually compliment those of the leading proponents of free culture and copyright reform. We do not accept that it is a zero sum game in which the efforts of one destroys that of the other. However, if you wish to focus on the role of a public advocate, it would be prudent to join them in their open rejection of our methods and limit any other guerilla activities.

Another kind of advocate is less public: to promote guerilla OA as well as legal OA among your friends, colleagues, and those who have the skills to be of use to the movement. Of particular importance are efforts to convert the casual pirate into the open access guerilla; from someone who copies content only for their own consumption to someone who recognizes the value of a more active and altruistic participation in the movement.

**The Prospector**

The role of Prospector is that of the scout for the movement. A Prospector identifies databases or collections of interest to the movement and collects information about its workings. What kind and how much content is there? How is it organized? What metadata is provided? What is the URL structure for the database? What is required for access and who provides the service? And so on.

We need to design a good system for compiling and sharing information of this nature among us, so that that the Armoror, the Sapper, or the Traitor can do their work.

**The Scribe**

The Scribe is a unique role. Scholars and collectors of all kinds have massive collections of material obtained by photographing, scanning, or

transcribing documents or assembling other digital assets. The resulting digitized content often sits on their own hard drives and are used in only one or a few publications or exhibits.

A Scribe in the movement is conscious of the importance of their videos, images, sounds, and digital archive photos.  When reasonable, the Scribe collects or takes photos that go beyond their own limited interest, or transcribes or indexes materials that may be of interest to others. They organize their information to the extent possible and make efforts to share their files widely. Materials not protected by copyright are to made public, directly posted online as public domain. When materials are suspected of being protected by copyright, they are distributed through other means or deposited with a Custodian. The Scribe is one of the roles that must take particular note of the responsibilities that go along with the safety and privacy of individuals and communities affected by the contents of the materials they digitize.

To facilitate the full integration of the Scribe into the movement, we must work towards better systems of making these collections easy to share, and a standard for describing and organizing them.

**The Courier**

The Courier is usually someone who has received collections of materials from another OA guerilla. They do not merely use the materials themselves, but recognize their obligation to help further share and distribute the materials widely.

While taking measures to protect themselves, the Courier makes efforts to share the materials online through torrents or other private repositories and servers, possibly coordinating these efforts with a Custodian. They share copies of the collections on portable storage media throughout their personal networks.

Another form of Courier plays the role of communication mediater between guerillas that should work to be kept in isolation from eachother, such as between the Traitor and Prospector and the Armorer.

**The Innkeeper**

The role of the Innkeeper is to manage the safe houses of our movement. We need safe and secure places to communicate with eachother anonymously. Ideally, these places should be kept isolated from any servers operated by the Custodian so that discovery of one does not compromise the other. The Innkeeper may be willing to host the work produced by the Armorer, various versions and updates of this cookbook, and other instructional materials.

An Innkeeper must be willing to maintain a communication network that will likely come under attack from hired hackers, botnets, or directly by principled opponents. They must have precautions in place to destroy anything that might betray the identity of our members. They should have plans to rapidly reproduce the network at a new location when taken down. They must lead efforts to detect moles and informants within the network and deny them access.

**The Armorer**

The Armorer is one of the most important roles in our movement. They create, maintain, and supply our movement with the weapons we need to carry out our raids. They write and update the scrapers to liberate content. They create the processing scripts to organize our files, and they design the protective measures that conceal our efforts.

In the past, the open access guerilla has often been an Armorer, a Traitor, and Custodian all in one. One huge disadvantage to this is that the Armorer who is also the Traitor cannot easily share their tools without potentially

coming under scrutiny for launching raids themselves. If discovered, as Custodian, their liberated materials are potentially surrendered and lost.

We must work together. If an armorer works together with Prospectors to identify targets and design scrapers, but maintains some distance from (or at least communicates anonymously with) the Traitors who will deploy them, we will stand a better chance of a successful operation. The Armorer may choose to work openly, if they protect connections to others in the movement. Writing a scraper is not necessarily a crime, but it is strongly suggested that efforts are made to limit distribution to a trusted network, at least until an operation is complete. This will delay any countermeasures by content distributors. More broadly, however, Armorers should be willing to share, through work such as adding recipes to this document, tutorials on general approaches to scraping databases and archives.

**The Sapper**

The Sapper is a special kind of Armorer. The Armorer serves primarily the Traitor, who will deploy scrapers from within a paywall or behind restrictive terms of service. The Sapper explores ways to infiltrate the security of archives and databases and enable outsiders direct access. They may hack servers directly in order to enable a full and immediate grab of the databases within. They may create access tunnels for a more cautious silent raid from the outside. Or, in the most extreme case, they may bring about a temporary destruction of security to allow large numbers of users to storm the database in a mass action.

The role of Sapper requires the greatest amount of skill and assumes the greatest amount of risk, both to the Sapper and the movement as a whole. Sappers should carefully consider the consequences of their actions and the impact on the ecology of content creation, curation, and consumption. A Sapper's raid, depending on how it is carried out, can be a massive act of sabotage, and has the greatest potential to generate anger and fear from our opponents but also put pressure on our sympathizers. Use of it as a tactic should be carefully considered, and great care taken in selecting the target, the timing, and the approach.

**The Traitor**

The Traitor is at the heart of the content liberation effort. They have legitimate and legal access to content targetted for liberation and release what they take to the Custodians and Couriers. They often depend on their special access to carry out their own daily tasks, and beyond legal consequences, may sacrifice much if their actions are discovered and their access revoked. Beyond this risk, the traitor often has conflicted loyalties. They have received their access in trust, and by helping the guerilla open access movement, they are inevitably betraying that trust.

The Traitor may know especially well the great efforts required to fund, produce, curate, and host large collections of data. Even as they liberate content, they may be concerned that their actions will contravene the wishes and have some impact on people who may have only very reluctantly accepted the restrictive copyrights and terms of use that have been forced upon the product of their efforts by the institutions they serve. If you work with a Traitor, be sensitive to these concerns, and respectful of what may seem like arbitrary limitations they wish to place on the scale and nature of their cooperation with the movement.

As the one opening the gate, the risky work of the Traitor should ideally be carried out in total secrecy. They should communicate securely and anonymously with other members of the movement, and limit their other roles. They do not need the Armorer's technical skills if they can obtain (indirectly, through a Courier, or directly and anonymously from a movement resource) the necessary scrapers and other tools produced by the Armorers. It is important, however, for them to become familiar with security

measures to protect their identity and conceal their liberation efforts. They should also learn enough about the scrapers etc. that they use so that they can run them on their own computers and adhere to the basic scraper guidelines (below).

**The Custodian**

Traitors or Sappers who have liberated content should move as quickly as possible to deposit this content with Custodians. The role of the Custodian is first and foremost that of canonical preservation. They also play an important role as the primary distributor of content to Couriers. They should also keep themselves informed about the safest and most effective means to widely distribute content on file sharing networks, secret repositories, and by other means. When hosts have been raided; copies of content have been taken down by legal authorities; or hosted copies disappear through neglect (lack of seeders for torrents, etc.), it is the job of Custodians to take measures to get the content to new sources.

Whenever possible the Custodian should check that the collections they received to not bear traces of the origin, looking for signs of watermarked PDFs, scraper files with revealing login information, or other signs that would reveal the identity of the guerilla who liberated the content.

Custodians should take precautions against their own discovery, and arrange for copies of materials in their care to be deposited in a safe place should they be discovered. They should also ideally limit actions in other roles of the movement to limit the risk of exposure.

**The Archivist**

The role of the Archivist is to preserve and improve the integrity of liberated content. They identify missing or problematic metadata, they process and organize files, and potentially provide conversions of problematic formats that materials are found in. If independent operations liberate parts of collections, the Archivist can help bring them together. They create note documents to include in the distribution of liberated content which describes the scope of the collection, identifies problems in the material, provides information on the originating source, and suggests ways to cite it. They collaborate with Prospectors to identify further work that needs to be done on already raided collections, and with Custodians to spread the best possible version of a collection.

Archivists with strong digital skills can also create the tools and platforms, both local and hosted, that will allow users to conveniently search and browse liberated content. A zip file full of PDFs or movie files is an order of magnitute less useful than a collection which is well indexed, annotated, and conveniently searchable.

**The Sculptor**

The Sculptor is someone who is willing to use and create something new with liberated content. They analyze and study collections for use in their own work. They produce new works of art and culture. They remash, reproduce, and transform liberated content. They generate innovative ways for others to manipulate and use content.

Ideally we should all be Sculptors. The value of the guerilla open access movement comes from facilitating the Sculptors of today, tomorrow, and every day that content would otherwise be locked away behind copyright and restricted use.

Scraper Guidelines
---

Scrapers are scripts designed to selectively extract content from servers. They are often written in scripting languages such as Python, Ruby, or Perl

that can be run from a variety of operating systems. They are one of the most powerful tools of the guerilla open access movement.

When scrapers or other raiding tools are designed by the Armorer and deployed by the Traitor, the two general principles of *respect* and *concealment* should be followed. For this purpose keep these guidelines in mind:

1. Minimize disruption to the host and their other users by
2. Limiting the scale and speed of a raid appropriately and
3. Employing methods to mask your raid as reasonable use of the target resource

More specifically:

- Never attempt to grab an entire database in the space of hours or over a very short time span relative to its total size
- Generally limit the practice of multiple simultaneous downloads from a server.
- If a small number of simultaneous downloads are made, carry it out from multiple network locations and ideally with multiple access credentials.
- Do not place others at risk by using their network access credentials unless they understand the potential consequences and volunteer.
- Employ proxies, MAC spoofing, and other measures as needed to conceal access locations when this is possible.
- Use random intervals between downloads to simulate human behavior
- Limit the operation of scrapers to certain hours to simulate human behavior or bury activity in periods of large regular traffic
- Design scrapers to download materials randomly (while logging completed downloads) rather than in sequence, or else a random groups of smaller sequences consistent with human behavior

**Recipes**
---

The remainder of this cookbook should be composed of recipes. These may include instructions for the use of or the code for scrapers and other tools that are appropriate for wider distribution. They may include tutorials and descriptions of good practices for the various roles of the movement. They may describe appropriate security measures. They may recount past victories and failures of the movement–but in a way that does not compromise anyone's identity. They should ideally not include any direct links to online resources, but may provide suggestions on how to use search techniques to find them, as they may often move. Under each recipe, include the date it was written, an optional author pseudonym, and if you distribute an edited version of this document, update the timestamp and version information at the bottom for the cookbook as a whole.

Securing Communication
---

Security in communication between members of the movement is of great importance. While legal authorities will want to investigate our actions, perhaps of a greater threat is that publishers and content industries that have significant resources will want to undercover who we are and can easily outsource their work to hired hackers.

It is important to follow some basic guidelines that can be grouped as follows:

1. Create barriers of separation between your regular activity and movement activity
2. Mask your identity
3. Mask your location
4. Encrypt your information
5. Limited Circles of Trust

**Separate your Movement Activity**

The first principle is a general one that you should always keep in mind. Whenever possible, create a separate sphere for all things concerned with your activity in the movement. Some of these things are just common sense. If you email about the movement, do so from a separate email account, not your regular personal email account. If you tweet about the movement, then unless you are only an Advocate who is not connected with illegal operations of the movement, tweet from a separate account. If you write about the movement on a website or other online service (again unless you are only an Advocate) then do so from accounts set up for the purpose.

If you need to set up online service accounts that require email verification use email addresses that come from an online email service which does not require you to provide further means of identification. This secures your anonymity as long as you mask your location.

When you are working on something related to the movement, even if you have masked your location, avoid doing other things online that may associate your IP address with other activity online and thus make it possible to trace back to you.

Consider using a different browser for all your movement activity. Or, at the very least, use "privacy" or "secure" or "icognito" mode in your browser. This will prevent your browsing history from being saved. More importantly it will prevent cookies from operation. If you are logged into a social networking service and then start doing movement activity without this layer of security in another tab or window, the website you are using may have ways to identify you through cookies.

**Mask Your Identity**

If you are a Courier or an Advocate, what you write and do will be exposed to the scrutiny of those outside the movement. Adopt a code name, but be sure to choose one that cannot be associated with you, even by friends. Thus, if you are a famous dog trainer, don't choose a code name of a breed of dog.

What you write can be subject to automatic text analysis. Authorship can be compared by means of algorithms. Try to vary your writing style. Make a list of idiosyncratic adjectives or other turns of phrase that you only use in some texts you write but not others. Write verbosely in some places, and in a short blunt manner in another. Use leet speak in some contexts, and grammatically correct language elsewhere. If you are exposed, this will help prevent you from being associated with all of your activity.

**Mask Your Location**

This is very important. Do not engage in movement activity from your own IP address (your identifying address on the internet) and spoof your MAC (the hardware address on your network or wireless card). If you are connected to a University of Vienna computer network, and you do anything online without masking your location, it is easy to trace the activity back to the university network, which will likely have a log of who is registered to use the IP, or at least the rough physical location that the person connected from. From there it is either direct discovery or discovery with subsequent surveillance.

*Mask your IP*

Use a free or paid VPN (virtual private network) service that cannot be traced back to you personally. Make sure to configure it so that all traffic to and from your computer is routed through this VPN. Connected to a VPN in Russia, a user on the network in Vienna will appear to be connecting from Russia.

*Spoof your MAC*

This is easily done with a little bit of experience on the command line. There are many utilities that can help you do this on Windows. On Linux or OS X simply open a terminal and enter the appropriate command that you can find many places online.

*Fool Your Enemy*

If, as a Courier for example, you engage in movement activity that may give indirect hints about your geographical location, then cloud their view. If you have a twitter account and are based in Canada, follow a set of users that might imply you are in France and read French. If you are emailing users in the UK, consider doing it from a free email service based in Germany.

**Encrypt Your Communication**

All movement related activity through email and chat should be secure or anonymous and generally both.

*Email and General Encryption*

Learn about GnuPG and the basics of public-private key encryption. Create an encryption key for yourself. Make sure the passphrase is very long. "The yellow elephant flew effortlessly behind the old barn" is far more secure than "&%tX90!" and easier to remember. Export the public key as ASCII armored and give it to your movement contacts. They will use *your* public key to encrypt email they send to you. You will use *your* corresponding secret key to decrypt the email they sent you that was encrypted by *your* public key. The process is reversed when you email them. You will need *their* public key to email them. Sign your communications with your key to help confirm your identity.

*Secure Chat*

Consider using anonymous communication on a pre-agreed IRC chat for something that is simple and fast and can be performed directly in your (secure/icognito) browser (while connected to VPN). There are also many plug-ins to secure communication on popular chat protocols, as well as some dedicated secure chat clients. Make sure that neither you nor the person you are speaking with have the client configured to log the communication.

*Keep Movement Related Files Secure*

If you are apprehended all your computers will be taken. You cannot be coerced into providing passwords though, so if you are smart, you will keep movement files completely secure.

You can encrypt your files and directories directly with GnuPG. Another common method is to create an ecrypted partition or "disk image" that is encrypted. Boot this partition or disk image when you need access to your movement files. Or have a separate encrypted hard drive that you load when you are ready to do work for the movement.

**Limited Circles of Trust**

Do not tell all your cool friends that you are active in the movement. Limit this information to people you trust *and* who you believe can be recruited to active participation in the movement. Whenever possible, maintain a segregated guerilla cell structure that the movement has operated under up until now. If you recruit several people to work with you do not pass on information about their identity or contact info to the person who recruited you. If you have contact with Couriers, let your up-contact know of their existence so tasks to/from Courier can be transmitted

via you. One exception are competent Armorers. Their skills are in high demand, and if you find them, consider keeping them separate from your own cell and "passing them up" the chain to the person who connected you put them in contact with an Armorer or Innkeeper you know of. The scripts/tools they create should be distributed through the network securely and then at some point appropriate openly. As far as this author is aware, we have no centralized structure. We operate in loose and a cell based movement, with anonymous cross-cell communication in online forums and chat maintained by admins (Innkeepers) who don't usually know the identity of anyone on the platforms they manage. Knowledge of personal real identities should, whenever possible, *not* be available beyond the individual cell. Traitors (and the more rare Sapper) are to be protected at all costs since they are directly liable for legal prosectution.

If there are people who you personally know but do not trust, yet who you think can be recruited, use a Courier as an intermediary. The Courier establishes contact, and if it goes well, they can "hand back" or "pass on" the contact once secure anonymous means of communication have been established and a role identified.

Cells that personally know eachother should take steps to ensure that keys were exchanged in a way to guarantee their genuine nature.

Limiting the circle of trust is key to the survival of the movment. In the past, many online hacker networks have been broken when one member is discovered and agrees to cooperate in exchange for a lighter sentence. A previously trusted person will unwillingly do all they can to expose other cell members and escape their own punishment. Be suspicious of suddent attempts to get more personal information from you by contacts in the movement and again, by keeping the circle of trust small, damage can be contained.

*Version (date-number-author)*:
2013.1.16-1.0.1-yellowElephant
*Modified from:*
2012.1.13-1.0-williwaw



Changelog
---

1.0 williwaw - original posted

1.0.1 yellowElephant - added recommended security precautions for movement members