Biosecurity Library: General methods documentation

DRAFT Last updated on March 23, 2022

Overview

The Biosecurity Library aims to widely disseminate and share knowledge to help advance biosafety and biosecurity. Researchers conducted a series of interviews and convened a working group with a cross-section of biosecurity experts to identify and collate the set of tools presented in this library. The library is a searchable and filterable database designed to enable ready access to biosafety and biosecurity resources from around the globe, published by governmental, international, and non-governmental organizations. The library will continue to be updated as additional resources are identified. Please contact us with any questions or additions at biosecuritycentral@georgetown.edu.

As of March 2022, the library includes resources from international, regional, national and institutional-level sources dated from 1972-2022 and organizations authoring these resources represent over 25 countries. Resources are categorized as: 1. Policies and legislation, 2. Requirements and protocols, 3. Core guidance and recommendations, and 4. Training materials. Each resource is also annotated with more than two dozen metadata tags, including a long description designed to give important background information and annotation that captures linkages where one resource mentions another (see citation web generation documentation below).

This work dataset are available for use under the Creative Commons Attribution 4.0 International Public License (https://creativecommons.org/licenses/by/4.0/), with appropriate reference and acknowledgement of the original research teams, as listed on this site.

Data Collection

The resources included in this library are designed as a curated, expert-validated compilation of those actively in use by those working in the field of biosecurity and related fields. To that end, resources for this library were identified through a series of interviews with global biosecurity leaders to understand what biosecurity resources and tools exist. Over 35 individuals from 20 countries were interviewed over a six month period and three expert working group meetings were conducted to identify resources and tools and to understand their use and nature in tag library entries that surface for what a resource is used, by whom, and for what purposes.

After a resource was identified through an interview or by the working group and screened by a member of our team for relevance, a researcher coded the resource according to the method detailed below. Resources that focused on human health and safety in the laboratory setting were prioritized, although some animal health, bioterrorism and disease specific resources are included. To be included, the requirement was that a resource be recommended by multiple interviewees and/or working group members OR that the resource meet a unique need in the field (e.g., tailored to resource constrained settings) if recommended only once.

In addition to interviews, a request for resources was solicited from the international biosafety community, working with the International Federation of Biosafety Associations (IFBA). This request led to feedback from 15 experts in 13 countries, 7 of which were previously unrepresented from the interviews.

The working group was given two opportunities to provide feedback on the set of resources included, and many of the resource entries were reviewed by their authoring organization for accuracy. All feedback was reviewed, adjudicated, and the final inventory dataset reflects this feedback.

Data coding

The team developed a custom data taxonomy and data dictionary to define key metadata and organize the dataset. The overall dataset, including the metadata coded to describe each resource, was reviewed by the research team, the working group, and where possible by contacts within the authoring organization of the resource. A custom citations mapping generator was run on the data to identify relationships between resources, as described below. The Glossary below provides the guidance that researchers used to code the data.

Citations web

The resources in this library, which set standards and provide guidance for the implementation and development of biosecurity, exist not in a vacuum but in an interconnected web. For example, checklists cite the standards documents they implement, and the cornerstone policies of the field are often cited by other resources and each other. Visualizing the relationships between biosecurity resources helps biosecurity managers, trainers, laboratory staff, legislative leaders, and others better use and understand information by clearly showing what related resources may be relevant to them.

Unlike in the broader academic literature, biosecurity resources do not necessarily use a standardized approach for citations. For example, a single biosecurity resource may cite others using a mixture of citation styles. A resource may provide citations in footnotes or endnotes rather than in a bibliography. Additionally, a resource may cite another in its main text and by name only, e.g., "Biological Weapons Convention", rather than providing a full citation. Thus, we developed a customized and flexible method to identify the references from one resource to another within the contents of document text.

A partly automated text-matching method was used to generate citation webs from a repository of document files and URLs as it supported the requirements for: (1) parsing multiple resource formats, (2) detecting citations regardless of format or style used, (3) detecting citations in any location in a document, and (4) operating on large volumes of text relatively quickly compared to manual review.

The following algorithm was used to determine whether one biosecurity resource is cited by another biosecurity resource:

Algorithm to determine if Resource A cites Resource B

1. Scrape main text of Resource A from any of the following sources, as available:
    a. Document file (e.g., PDF)
    b. Website URL(s)
2. Generate query terms for Resource B based on its title
3. Search for Resource B's query terms within main text of Resource A, using exact matching or fuzzy matching (see below)
4. If at least one query term is found in Resource B, mark Resource A as citing Resource B
5. Manually review bulk results in the format of a network map to ensure detected citations are reasonable and match expectations based on knowledge of library

Query terms

Query terms are text strings, such as the name of a resource, that are searched for within the main text of another resource. If a query term is found in a resource's main text, it is considered to cite the resource whose query term it is.

For example, the resource with name "Laboratory Biosafety Manual Fourth Edition and Associated Monographs" has a set of query terms including its name and the following manually-specified alternate names that are known to be frequently used:

- Laboratory Biosafety Manual, Fourth Edition
- Laboratory Biosafety Manual, 4th Edition
- Laboratory biosafety manual, 4th edition
- Laboratory biosafety manual, fourth edition and associated monographs

Exact vs. fuzzy text matching

Mentions of resources are detected using either exact or fuzzy text matching depending on the conditions described below. In exact text matching, only the exact characters queried are considered a match. On the other hand, in fuzzy text matching, some imprecision with respect to spelling and case in the match is tolerated. For example, "Laboratory Biosafety Manual, Fourth Edition" almost exactly matches the phrase "Laboratory biosafety manual (Fourth Ed.)", but not quite, and the fuzzy text matching algorithm would report the partial match and a score quantifying its exactness.

Exact text matching is used when any of the following are true:

1. The query text is not at least six[1] words long
2. Resource name aliases are being searched for in addition to the resource's main name (i.e., title)[2]
3. The resource name is very similar to another resource in the library's name

If none of the above are true then fuzzy text matching is used in order to increase the quality of results. For example, the name "Laboratory Biosafety Manual, 4th Edition" is extremely similar to the name "Laboratory Biosafety Manual, 3rd Edition", which means the potential for false positives is high. This risk is mitigated with exact matching, used in this case because the query term is fewer than six words. On the other hand, the name "Africa CDC Biosafety and Biosecurity Initiative Report on the Consultative Process to Identify Priorities for Strengthening Biosafety and Biosecurity" is fairly long and likely to appear with a greater number of unpredictable spellings and capitalizations, so fuzzy matching is used.

In this application, when fuzzy matching is used, the match is required to be at least 90%[3] exact.

---

[1] This minimum word count was chosen heuristically based on inspection of typical results
[2] It is assumed that if resource name aliases are defined that fuzzy text matching should not be used because most expected spellings of the resource name will be covered by the resource name aliases and the resource's name.
[3] This minimum similarity level was chosen heuristically based on inspection of typical results

Citations and data use

Information from this project is openly available for use with attribution. Please use the following citation to acknowledge the work:

Georgetown University Center for Global Health Science & Security. Biosecurity Central. Washington, DC: Georgetown University. Available at https://biosecuritycentral.org/

We ask that you let us know if you publish anything using data from this site, so we can link to it. Contact us at biosecuritycentral@georgetown.edu

Glossary

This glossary is designed to provide the user with key information to understand the way items were coded and the meaning ascribed to them by the research team. Other explanations are available throughout the above methodological discussion, and through the tooltips available throughout the website.

Resource type

Describes the main content in a resource. Can be one of four options, detailed below.

Policies and legislation

Resources that are policies, legislation, international agreements, national implementation of international agreements, or other legally binding frameworks. Or, resources that deal exclusively with policies and legislations, but are not in themselves policy documents, such as reports and databases.

Training materials

Resources whose primary goal is to educate their audience on a biosecurity topic. The format of these resources are often webinars, courses, and certifications.

Requirements and protocols

Resources that outline requirements and criteria necessary to achieve certain certifications and standards, both legislative and norms-based consensus documents fall into this category. Also action plans and protocols to run a safe and secure lab or other similar institutions.

Core guidance and recommendations

Resources that do not have any binding authority, rather are best practices for maintaining biosecurity.

Key topic area

The main topic(s) that the resource aims to cover. Often, categories overlap and a resource will fall under multiple topic areas.

Risk assessment

Resources that categorize different levels of biological risk, or provide guidance on biorisk management.

Laboratory biosafety

Resources that outline how to run a safe laboratory, often coming as a joint resource on how to run a safe and secure lab.

Export controls

Resources that deal with which biological materials can be moved across borders, and how to be sure that these materials are moved in a safe, secure and mutually agreed upon manner.

Medical diagnostics

Resources that cover how to set-up and execute safe and reliable test results for infectious diseases; medical diagnostics are an important part of disease response.

Environmental safety

Resources with special considerations for Genetically Modified Organisms (GMOs).

Laboratory research

Resources that deal directly with the day-to-day operations of a research laboratory, or that discuss responsible research conduct.

Sample transportation

Resources that outline how to move biological samples from one location to the other. Guidelines could be national or international, and could be actual guidelines or frameworks for how to create transportation agreements between 2 entities.

Law enforcement

Resources for law enforcement that usually cover how to respond to a biological threat or incident. Resources cover preparedness and training through response and post-event investigations. Also covers coordination with other agencies and departments.

Zoonotic diseases

Resources that cover working with and preventing the spread of zoonotic disease. These resources often cover how to properly work with animals to minimize chance of exposure and spreading of zoonotic diseases.

High-consequence pathogens

Resources that deal with highly infectious diseases, both human and animal pathogens. These resources also usually deals with their containment

Animal health

Resources that are often aimed at veterinarians and farmers, but could also be aimed at labs that work closely with animals. These resources cover both how to keep animals healthy and how to keep sick animals contained so they do not spread their illness to other animals or humans.

Legal mechanisms and authorities

Resources that are policies, legislation, international agreements, national implementation of international agreements, or other legally binding mechanisms. Or, resources that deal with policies and legislations, but are not in themselves policy documents, such as reports and databases, as well as training materials.

Laboratory biosecurity

Resources that deal specifically with biosecurity in a laboratory setting.

Dual Use

Resources that address dual use research of concern (DURC) and the responsible conduct of researchers in relation to dual use.

Target and potential user role

The target user role is the user(s) listed by the resource as being the primary users. The potential user role is user(s) our team has identified as being interested in this resource, if applicable.

Legislative leader

Users that are responsible for making policy and legislation decisions at the national, regional or international level

Laboratory manager

Users that are responsible for the operations and management of a laboratory. These users set the standards by which a lab is run.

Laboratory technician/staff

Users that carry out the day-to-day technical work of a laboratory. Typically, these are the users that engage the most directly with potentially hazardous materials.

Military

Users that are members of national armed forces, that are often tasked with carrying out security related programs

Trainers and teachers

Users that train and teach in any capacity- biosafety trainers, university professors, high school teachers, etc.

Biosafety managers

Users that are responsible for ensuring biosafety, often at an institutional or government level. Very often this role involves setting policies and training individuals.

Law enforcement

Users that are members of law enforcement agencies at the local, national and international level.

Veterinarian/Farmer

Users that are responsible for the well-being of animals and for ensuring the biosafety and biosecurity of animal facilities.

Funder

Users that are interested in funding biosecurity-related projects and programs

Public Health Official

Users that are members of public health agencies at the local, national or international level

Government Official

Users that are members of other, non-public health, agencies at the local, national or international level

User type

The type(s) of users that would benefit from using the resource.

International

These resources are for users who think at the international level and want to know about standards, guidelines and treaties that impact the world.

Regional

These resources are for users who are interested in initiatives, networks and policies that are being formed and enforced in a regional setting.

National

These resources are for users who are looking for national level guidance and implementation policies and strategies for biosecurity. Many of these resources can be used as case studies for other nations.

Institutional

These resources are for users who are looking for lab level guidance, rules, and trainings.

Authoring organization type

The type of organization who authored the resource.

International

Organizations that have international authority

Regional

Organizations that have regional authority

Bilateral

When a resource is the result of two or more organizations from different countries collaborating and building off of each other's expertise and authority.

Government

Organizations that are part of their national government and have legal authority

National

Organizations that have national authority, but are not part of the government and do not have legal authority.

Foundation/non-profit

Private organizations that are not for profit

Academic

Private organizations that are associated with an academic institution

Private Sector

Private organizations that are for profit