ESCAP
Institutional Repository Policies

Contents

About this document

This document is part of the Institutional Repository project, in order publishing different contents according to the international standards and being part of the international guidelines.

Version history

VERSIÓN	APPROVAL DATE	WRITTEN BY	REVIEWED BY	APPROVED BY	MODIFICATION DESCRIPTION
1.0	feb-04-2021	Joel Torres			The first version of the reference model is created.
1.1	Sep-03-2021	Benjade Palmero			The updated version after the final migration.

Collections and Communities management

According to the definitions and specifications given by the ISO 16363:2012 standard, storing, organizing and ensuring the preservation of information in the long term is a most-activity within the functions of a repository.

To fulfill these functions, it is essential that the repository has a structure of communities and collections that allows it to have content segmentation according to the groups that are in charge of managing them, as well as a clear organization that allows them to be grouped by collections according to its typology.

The repository has been implemented on the DSpace technological platform which, from its conceptualization, is focused on receiving content, facilitating its description, storing it, facilitating its location and ensuring its preservation.

With regard to the organization of content, the digital repository operates from four fundamental concepts:

File: refers to the digital files that will be delivered to the repository in various technical formats (e.g. TXT, CSV, PDF, XML, JPG) and that make up an intellectual entity (e.g. article, chapter, book, working paper). An intellectual entity can be made up of one or more files; For instance, a journal article represents for a reader a single intellectual entity, which technically can be stored in a single PDF file, or in several XML, JPG and MP3 files.

Metadata: are the elements that will be used to describe the content, the technical format and the information related to the intellectual property and rights of the content deposited in the repository. The use of metadata is described in detail later in this document.

Item: it is known as the conformation of an information package that includes all the files and all the metadata that shape an intellectual entity and that will be managed by the repository to carry out the tasks of storage, preservation, exchange or presentation.

Collection: a collection within the repository is made up of a group of items that share similar characteristics due to their typology. For instance, there may be a single collection of working papers, a collection of books, a collection of articles, and so on.

Community: it is the grouping of several collections according to their origin. In other words, collections can be grouped according to the way in which people produce and deliver information within the organization.
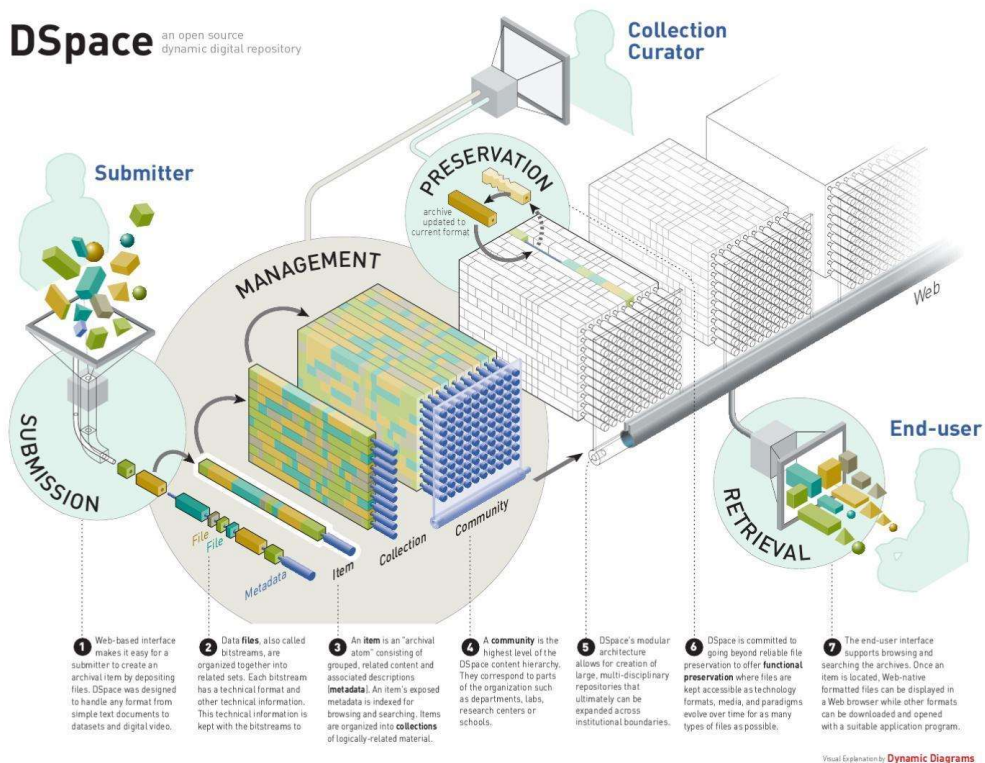
Figure 1: Files, metadata, items, collections and communities structure on DSpace.

Source: https://duraspace.org/dspace/about/features/

Communities and Collections definitions

Name of the community	Community's collections
ESCAP Publications	❖ **Journals** ➢ **Asia-Pacific Development Journal** ➢ **Asia-Pacific Population Journal** ➢ **Asia-Pacific Sustainable Development Journal** ➢ **Asia-Pacific Tech Monitor** ➢ **Asia-Pacific Trade and Investment Review** ➢ **Environment and Sustainable Development News** ➢ **Flood Control Journal**

	➢ **Industrial and Technological Development News for Asia and the Pacific** ➢ **Population Headliners** ➢ **Population Research Leads** ➢ **Poverty Alleviation Initiatives** ➢ **Statistical Indicators for Asia and the Pacific** ➢ **Statistical Newsletter** ➢ **Transport and Communications Bulletin for Asia and the Pacific** ➢ **VATIS Updates** ➢ **Water Resources Journal** ❖ **Flagships** ❖ **Manuals & Training materials** ❖ **Maps** ❖ **Policy briefs** ❖ **Reports** ❖ **Working papers**

Metadata schema

Among the main functions of the repository is to strengthen the visibility of the information, making it necessary to ensure the delivery of its metadata to various types of platforms in a structured and standardized way (interoperability).

This interoperability is based on three main bases: the information exchange protocol, the metadata schema and the syntax guidelines that will be applied to the metadata.

For the information exchange protocol, the repository will adhere to the specifications of the standard known as OAI-PMH; Dublin Core schema will be used for the metadata schema, and OpenAIRE recommendations will be applied for the textual syntax, which will be complemented with specifications that are described below.

About Dublin Core

In 1995 in the city of Dublin, Ohio, from a call for work attended by experts in various fields such as library science, text marking, digital publishing and content distribution, the Dublin Core Metadata Initiative was created, with the aim of creating a set of descriptors that were easy to follow and understand, also allowing the organization and retrieval of various types of content through a common vocabulary, so that users can locate the content even if they are from different disciplines.

Currently, the Dublin Core scheme has been translated into more than 20 languages, counting officially in the WWW Consortium. In addition to having been adopted as an official standard in the USA through the ANSI / NISO z39.85 standard and in Europe by the CEN / ISSS (European Committee for Standardization / Standardization System for the Information Society) through the ISO 23950 standard. It can also be found in two Requests For Comments: RFC 5013 and RFC 5791.

For a better understanding, it is necessary to list some important definitions:

DCMI

It is the acronym for the Dublin Core Metadata Initiative and it is the organization in charge of maintaining the Dublin Core metadata standard.

Resource

It is the content to publish and preserve, according to the definitions made previously in this guideline, regardless of its typology and format.

Life cycle of a resource

It is a sequence of events that mark the development, transformation and use of a resource, from its creation, its processes of review, approval, publication, deposit, consultation, to the necessary activities for the preservation of the content in the long term, including any alteration that is made, either by systems or by people.

Element

It is each of the descriptive labels that, using a common semantics, seek to facilitate the understanding for users of the characteristics, content, format and rights of the resource.

Set of elements

They are all descriptive labels that together allow the resource to be correctly described. To do this, each tag uses a unique word, which is understandable to users and can be interpreted by computers to create encoding schemes and properly process the information.

These tags are grouped by type, according to what describe about:

- Those for content, formats and technical aspects,
- Those for information related to intellectual property and rights.

Second, the labels have two levels: the basic one, made up of a set of 15 elements, and the qualified one, made up of sub-elements that can be related to the basic elements.

General recommendations for using Dublin Core

Although certain languages allow the indistinct use of capital letters and lowercases in the definition of labels, it is best to adhere to the specification respecting the use of them in the names of the elements, this will allow greater interoperability in cases where information is interconnected with systems or languages such as XML, where it is a case-sensitive.

Each element is optional and can be repeated according to the needs of the organization, however, its use must adhere to the instructions indicated later in this document.

More information

For more information regarding the use of Dublin Core, controlled vocabularies or details of the specification, it is necessary to consult the official site: https://www.dublincore.org

About OpenAIRE

It is a technological and services infrastructure created in 2009 to support, accelerate and measure the correct implementation of European policies for open access to scientific publications and research data. OpenAIRE has a strong network of National Open Access Desks (NOADs), agents that act as national reference points to disseminate the open access policies of the European Commission among institutions and researchers, as well as to facilitate the coordination of national policies with European ones.

Within its services, OpenAIRE develops and publishes guidelines for the use of metadata based on Dublin Core for the description of textual resources, multimedia material, data sets, software, patents, books, working papers, among others.

Adherence to the guidelines published by OpenAIRE, allows to ensure the interoperability of the Digital Repository with other repositories, searchers, indexes, harvesters and other types of specialized systems around the world that facilitate the location of content by the public.

More information

For textual resources, adherence to the specifications described in OpenAIRE Guidelines for Literature Repositories in its v3 version is recommended, in particular, the use of those elements defined as mandatory (M) is recommended:https://guidelines.openaire.eu/en/latest/literature/index_guidelines-lit_v3.html

For data set or software type resources, adherence to the specifications described in OpenAIRE Guidelines for Data Archives version v2 is recommended, in particular, the use of those elements defined as mandatory (M) is recommended: https://guidelines.openaire.eu/en/latest/data/introduction.html

About the metadata

It is necessary to be guided by the next table:

Title (M)	**Enter the main title, including subtitle, of the item here. Separate subtitle from main title with a colon (:). Capitalize the first letter of the first word only, except for proper nouns (person, place, or organization).**	**dc.title**
Other Titles (MA)	**Enter any alternative title here, for example a title in another language, a common acronym for a journal title, etc...**	**dc.title.alternative**
Corporate author(s)	**Enter the organization(s) or corporate author(s) which are responsible for the creation of the item.**	**dc.creator**
Authors (M)	**Enter the personal author(s) responsible for creating the resource.**	**dc.contributor.author**

Editor(s) (MA)	Enter person(s) responsible for editing the resource.	dc.contributor.editor
Other Contributor(s)	Enter other contributor(s) to the resource.	dc.contributor.other
Summary / Abstract (MA)	Enter a summary or abstract for this item.	**dc.description.abstract**
Summary HTML/ Abstract HTML	Enter a summary or abstract for this item in HTML	**dc.description.html**
Keywords ®	Enter any keywords not currently available in the UNBIS Thesaurus. Please make every effort to find a relevant UNBIS term rather than enter keywords here.	dc.subject
Resource type / Content type (M)	Select the type of content from the list below. Tip: Select "Text" unless otherwise instructed.	**dc.type**
Issue date / Publication date (M)	Enter date issued as printed in publication. You can leave out the day and/or month if they aren't applicable.	**dc.date.issued**
Language (MA)	Select the language of the main content of the item. If the language does not appear in the list, please select 'Other'. If the content does not really have a language (for example, if it is a dataset or an image) please select 'N/A'.	dc.language.iso
Publisher (MA)	Enter the name of the publisher. For official publications, this will usually be "United Nations" unless otherwise noted on inside cover publication information. If no publication information is provided, you may use "UN. ESCAP".	dc.publisher
Place of publication	Enter the location/place where the item was published.	escap.publisherPlace
Identifiers ®	Select the identifier from the list below and enter the identification number in the corresponding field.	dc.identifier
ESCAP Document type / Publication Type (M)	Select the type of publication or knowledge product from the provided list.	escap.doctype
Table of contents (MA)	Enter the Table of Contents (if available).	dc.description.tableofcontents

Description (MA)	Enter additional information about the publication here not covered by the Abstract or Summary or other descriptive fields. This might be a theme, background information, etc...	**dc.description**
Copyrighted date (MA)	Enter copyright date of the resource.	dc.date.copyright
Extent / Pages ®	Enter the number of pages and other available descriptive format elements (illustrations, maps, etc...)	dc.format.extent
Bibliographic citation ®	Enter the standard citation for the previously issued instance of this item.	dc.identifier.citation
Series/Journal Title (MA)	Enter the series or journal title as printed on the cover or first page. As you start typing, you will find a drop down list that will help you find existing approved entries.	dc.relation.ispartofseries
Series/Volume Issue	This field is used to identify and group volumes and issues of the same series. Enter the same series title as above, and in the second box, enter the Volume and Issue information in the following format: Vol.4, no.23 (or similar.)	escap.ispartofseriesno
Access rights (MA)	Information about who accesses the resource or an indication of its security status. Access Rights may include information regarding access or restrictions based on privacy, security, or other policies.	dcterm.accessRights
Audience (O)	Information or list of a group or a population sector to whom the resource is intended or useful.	dcterms.audience
ESCAP Programme of work (MA)	Select the ESCAP Programme of work from the list provided.	escap.programmeOfWork
ESCAP Areas of work (MA)	Select the ESCAP Area(s) of work from the list provided.	escap.areasOfWork
ESCAP Subregion	Enter the ESCAP Subregion(s) covered topically by the resource. Click in the box to select the relevant term(s) from the list provided. Please use the Country/Region field above for all other geographical descriptors.	escap.subregion
Subject(s) (MA)	Click in the box to select the relevant subject/keyword term(s) from the UNBIS Thesaurus.	dc.subject.unbist
UN Document symbol (MA)	Enter the UN Document Symbol here.	escap.unDocSymbol

UN Sales no. (MA)	Enter the UN Publications Sales Number here.	escap.unSalesNumber
ESCAP Bibliographic Record Number ®	Enter the ESCAP Bib # assigned in the Koha ILS (if applicable).	escap.libBibNumber
ESCAP Call Number	Enter the ESCAP Call Number (if applicable).	escap.libCallnumber
External Link(s) (O)	Enter the URL of ESCAP publications on located on ESCAP website (Drupal)	escap.link.externalLink
Internal Link(s)(O)	Enter the URL of ESCAP publications on the external website	escap.link.internalLink
Media type ®	Select the media type from the list below.	dc.format.mimetype
Right statement ®	ESCAP's current default rights statement is automatically entered. IF the publication is a co-publication or collaboration with additional or alternative rights statements, please edit the field as required.	dc.rights
Sponsorship	Enter information about sponsoring agencies, individuals, or contractual arrangements for the item.	dc.description.sponsorship
Time period ®	Enter the temporal coverage or Time period(s) covered.	dc.coverage.temporal
Country / Region ®	Enter the geographic area(s) covered topically (country, region, etc.) by the resource. Click in the box to select the relevant term(s) from the UNBIS Thesaurus.	dc.coverage.spatial
Right holder(s) (MA)	Please enter United Nations and any additional copyright holders (in the case of co-publications.)	dc.rights.holder
Journal Title ®	The title name of the container (e.g. journal, book, conference) this work is published in. This property is considered to be part of the bibliographic citation.	oaire.citationTitle
Volume no. ®	The volume, typically a number, of the container (e.g. journal). This property is considered to be part of the bibliographic citation.	oaire.citationVolume
Issue no. ®	The issue of the container (e.g. journal). This property is considered to be part of the bibliographic citation.	oaire.citationIssue

Article First Page no. (For citation) ®	**The start page is part of the pagination information of the work published in a container (e.g. journal issue). This property is considered to be part of the bibliographic citation.**	oaire.citationStartPage
Article End Page no. (For citation) ®	**The end page is part of the pagination information of the work published in a container (e.g. journal issue). This property is considered to be part of the bibliographic citation.**	oaire.citationEndPage
Edition ®	**The edition the work was published in (e.g. book edition). This property is considered to be part of the bibliographic citation.**	oaire.citationEdition
Bibliographic level	**Select the bibliographic level (monograph, serials, etc...) from the drop-down menu below.**	escap.bibLevel
UN SDG (MA)	**Select the SDG Goals from the list below.**	escap.unSDG
Project Title (MA)	**This field is used to identify and group volumes and issues of the same series. Enter the same series title as above, and in the second box, enter the Volume and Issue information in the following format: Vol.4, no.23 (or similar.)**	escap.projectName
Contact Unit	**Select the ESCAP's substantive divisions or subregional offices from the list provided.**	escap.contactUnit
Contact Phone	**Enter contact phone number eg +66 2 288-1234**	escap.contactPhone
Contact Email	**Enter contact email address**	escap.contactEmail
Event Name (MA)	**Enter meeting/event name**	escap.eventName
Event Type (MA)	**Select the meeting/event type from the drop-down menu below.**	escap.eventType
Event Date (MA)	**Enter date of meeting/event**	escap.eventDate
Event Location (MA)	**Enter location of meeting/event**	escap.eventLocation

M = Mandatory | MA = Mandatory if is applicable | R = Recommended | O = Optional

To know more details and consult the current specification, visit the next table of digital metadata, on:

https://docs.google.com/spreadsheets/d/1CX-xTy3MV8Pv_JAsboNPCaEmik5WwaP1GE2jd_iWmHA/edit#gid=1624786860
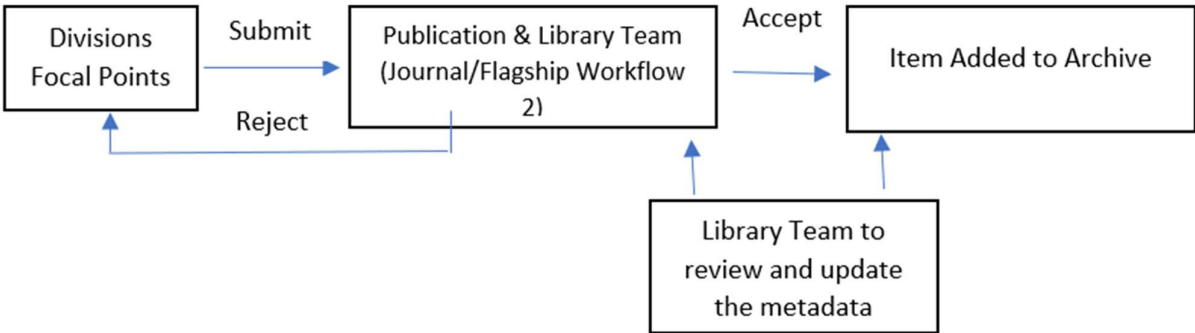
Workflow

In the ESCAP Repository, two submission workflows are identified for (a) Journals and Flagships, and (b) Manuals & Training materials, Maps, Policy briefs, Working papers, and Reports.

**(a) Workflow steps for Journals and Flagships**

The image below depicts the submission process if someone is assigned to each step.

The sequence is this:

The Divisions' focal point submits an item to the collection. The publication and Library team is notified of the submission by email. If the submission is rejected, the item is returned to the submitter and the submitter can modify and resubmit. If the submission is accepted, the item will be archived in the repository and available on the public site. The Library team will then review and edit the metadata of accepted item.

(b) Workflow steps for Manuals & Training Materials, Maps, Policy Briefs, Working Papers and Reports
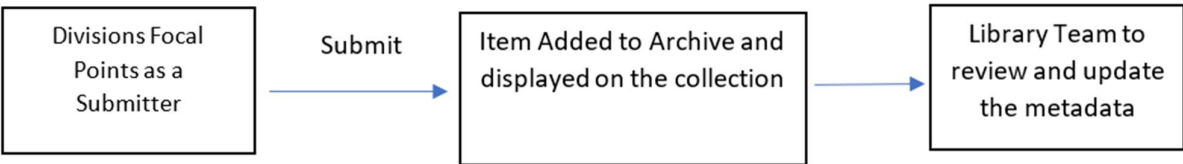
The sequence is this:

The Divisions' focal point submits an item to the collection. The item will be archived in the repository and immediately available on the public site. The Library team then will review and edit the metadata.

Metadata to present step-by-step

The system will show the next metadata for each workflow step:

Step 1	**Title (M)** **Corporate author(s)** **Authors (M)** **Editor(s)** **Other Contributor(s)** **Publisher (M)** **Place of publication** **Issue date / Publication date (M)** **Copyright date (MA)** **Identifiers ®** **UN Document Symbol (MA)** **UN Sales Number (MA)** **Series/Journal Title (MA)** **Series Volume/Issue (MA)** **Project Title (MA)** **ESCAP Document type / Publication Type (M)** **ESCAP Programme of work (MA)** **ESCAP Areas of work (MA)** **ESCAP Subregion** **Country / Region ®** **Subject(s) (M)** **Keywords ®** **Time period ®** **Summary / Abstract (MA)** **Table of contents (MA)** **Description (MA)** **Contact Email (MA)** **Contact Phone (MA)** **Contact Unit (MA)**
Step 2	**Resource type / Content type (M)** **Media type ®** **Extent / Pages ®** **Language (MA)** **UN SDG (MA)** **Sponsorship** **Bibliographic level** **ESCAP Bibliographic Record Number ®** **ESCAP Call Number** **Event Name (MA)** **Event Type (MA)** **Event Date (MA)** **Event Location (MA)**

Formats and technical characteristics of the contents

This section is made up of a recommendation series to follow for a better management of the contents regarding the technical formats in which they are stored in digital form. This aims to facilitate the processes of preservation, storage, presentation and exchange of information between systems in the long term.

The format is the structure and encoding of what a digital object is created with, and enables their interpretation and presentation by some computing device.

There may be different formats that allow the interpretation and presentation of different types of content, whether they are text documents, images, video, audio, data sets or other types of digital objects.

According to the Open Data Handbook (https://opendatahandbook.org), there are open formats and closed formats, which are defined below:

> *"An **open format** is the one where the software specifications are available to anyone, free of charge, so anyone can use those specifications in their own software without any limitation on their reuse that is imposed by intellectual property rights.*
>
> *If the file format is **'closed'**, this may be due to the format being proprietary and its specifications are not publicly available, or because the format is proprietary and although the specifications have been made public, its reuse is limited. "*

From this it follows the importance of the use of open formats for academic and scientific research, since it is the only way to ensure that the content can be processed correctly by the technologies of the future.

This document presents a general guide to the use of these types of formats.

Recommended standards for formats

It is important to start from the fundamental premise that any process related to the construction and publication of digital information must consider the use of open formats at the same time the content is created, therefore, it is very important to achieve, as far as possible, the participation and cooperation of the different actors, whether they are authors, reviewers, various collaborators, assistants, information managers, trainers and anyone who affects the content at any point in the process.

That is why information systems administrators must correctly evaluate and select the type of formats on which they will work, ensuring that they are open, highly recognized, flexible and with the ability to operate on various technologies.

General considerations to take into account when selecting formats:

1. Make a review of the most commonly used formats in the area of knowledge on which you are working, as well as in similar organizations.
2. Consider the type of interaction that is expected to be obtained regarding the content (if the content must be descriptive, if it must have the ability to be segmented, if it requires having some functions that involve performing calculations or functions, etc.).
3. Identify the range of time that content should be stored and kept accessible.
4. Have parallel versions and content review processes to ensure that, during future format conversion processes, there is no data loss.
5. Make periodic reviews on technological trends in relation to open formats.
6. They should not handle encryption or compression, as this will require third-party interpreters to use them.

Recomended standards

Consultation and adherence to the following standards, recommendations and good practices are highly recommended:

FAIR principles

The FAIR Guiding Principles for scientific data management and stewardship: https://www.nature.com/articles/sdata201618.

They are the guiding principles for making research-data easy to find, accessible, interoperable and reusable, and they provide a guide for managing scientific data that suggest any data set should have the structure that allows them to be:

1. **FINDABLE**, through the assignment of persistent identifiers, using standardized and detailed metadata, registering them in searchers or systems and the inclusion of persistent identifiers in the metadata that describe the data.
2. **ACCESSIBLE**, so they can be downloaded. It is necessary to use standardized, open, free and universally implemented communication protocols to recover them. This allows, if necessary, to maintain the record of metadata and manage user authentication and authorization of use.
3. **INTEROPERABLE**, through the use of a formal, accessible, shareable and widely applicable language, as well as the use of controlled vocabularies that, at the same time, are compatible with the FAIR principles, and the data and metadata include qualified references to other data or other metadata.
4. **REUSABLE**, allowing other people to reuse them, making clear their origin and the conditions of reuse. It is necessary they have precise and relevant attributes, with clear and accessible licenses, and a

clear origin, following the standards used by the community related to the field for which the data was created.

UTF-8 character encoding

Choosing and applying a character encoding, W3C internationalization recommendations (https://www.w3.org/International/questions/qa-choosing-encodings).

These guidelines should be taken into account by HTML coders (using editors or scripts), script developers (PHP, JSP, etc.), CSS coders, Web project managers, and anyone who is new about character encoding and needs an introduction to choosing and applying character encoding.

This practice allows the correct use and better control of languages in the content they present and on websites. The UTF-8 is also the most widespread encoding on the Internet.

Definition of open formats to use

For the use of formats, it is recommended to consider the following aspects:

❖ The selected formats must have an open, public specification and complete documentation.
❖ The formats should not have legal, technical or financial barriers that prevent or limit their free use, either to build files, interpret or transform them.
❖ Use file formats supported by FAIR (source: https://biblioguias.cepal.org/gestion-de-datos-de-investigacion/formatos):
  ➢ **Containers:** TAR, GZIP, ZIP
  ➢ **Databases:** XML, CSV, JSON
  ➢ **Geospatial:** SHP, DBF, GeoTIFF, NetCDF
  ➢ **Video:** MPEG, AVI, MXF, MKV
  ➢ **Sound:** WAVE, AIFF, MP3, MXF FLAC
  ➢ **Statistics:** DTA, POR, SAS, SAV
  ➢ **Images:** TIFF, JPEG 2000, PDF, DNG, GIF, BMP, SVG
  ➢ **Tabular data:** CSV, TXT
  ➢ **Text:** XML, PDF / A, HTML, JSON, TXT, RTF
  ➢ **Web Archive:** WARC

General table of formats to use and their level of openness:

Format	Brief description	Data type	Openness level
PDF	PDFs are text files that are not in structured format. They are used for document generation, not for publishing or storing data.	Text	very low
XLS	XLS are spreadsheet files. It is a proprietary Microsoft format.	Tabular	low
XLSX	The XLSX are also spreadsheet files whose format was developed by Microsoft but its specification is open (ISO / IEC 29500: 2008). It is the default format for Excel 2007 onwards.	Tabular	Medium
ODS	ODS are files with the structure of an XML. It is an open format based on the OASIS OpenDocument Format (ISO / IEC 26300). It is the default format of the Open Office spreadsheet processor.	Tabular	Medium
CSV	CSV files are plain text files where columns are separated by commas and rows by line breaks. It is an open format.	Tabular	High
JSON	It is a format for the exchange of data between systems. It is an open non-tabular format based on the RFC 7159 specification.	Structured	High
SHP	It is a format for the exchange of data between systems. It is an open non-tabular format based on the RFC 7159 specification.	Geographic	Medium
KML	It is an open format for geographic data based on the XML standard.	Geographic	High
GEOJSON	It is an open standard format designed to represent simple geographic features, along with their non-spatial attributes.	Geographic	High
GEOPACKAGE	It is a geospatial data format implemented as a SQLite database container.	Geographic	High

Source: Guide for the publication of data in open formats (https://datosgobar.github.io/paquete-apertura-datos/guia-abiertos/)

Definition of the structural schemes to follow

Next, the basic schemes to which the contents delivered in certain formats must be adhered to are presented.

For datasets in CSV format: it is necessary to adhere to the Common Format and MIME Type for Comma-Separated Values (CSV) Files specification, published by The Internet Society in 2005 (https://tools.ietf.org/html/rfc4180).

1. For the exchange of JSON data: it is necessary to adhere to the specification The JavaScript Object Notation (JSON) Data Interchange Format, published by The Internet Society in 2014 (https://tools.ietf.org/html/rfc7159).

2. For SVG vector graphics: take the guide from the description Scalable Vector Graphics (SVG), Version 1.1 published by the Library of Congress, updated online in March 2020 (https://www.loc.gov/preservation/digital/formats/fdd/fdd000020.shtml).

3. For TIFF images: take the guide from by the TIFF description, Revision 6.0, published by the Library of Congress, updated 2009 (https://www.loc.gov/preservation/digital/formats/fdd/fdd000022.shtml).

4. For WAV audio files: take the guide from the WAVE Audio File Format description, published by the Library of Congress, updated 2012 (https://www.loc.gov/preservation/digital/formats/fdd/fdd000001.shtml).

5. For MP3 audio files: base on the description MP3 File Format, published by the Library of Congress, updated 2010 (https://www.loc.gov/preservation/digital/formats/fdd/fdd000105.shtml).

6. For RIFF multimedia resource exchange files: be guided by the description RIFF Resource Interchange File Format, published by the Library of Congress, updated in September 2004 (https://www.loc.gov/preservation/digital/formats/fdd/fdd000025.shtml).

7. For MPEG video files: take the guide from the description MPEG-4 File Format, V2, published by the Library of Congress, updated in 2012 (https://www.loc.gov/preservation/digital/formats/fdd/fdd000155.shtml).

8. For AVI video files: be guided by the description AVI Audio Video Interleaved File Format, published by the Library of Congress, updated in March 2016 (https://www.loc.gov/preservation/digital/formats/fdd/fdd000059.shtml).

9. For MPEG video files: be guided by the description MPEG-4 File Format, V2, published by the Library of Congress, updated in 2012 (https://www.loc.gov/preservation/digital/formats/fdd/fdd000155.shtml).

10. For PDF text files: be guided by the description PDF / A Family, PDF for Long-term Preservation, published by the Library of Congress, updated December 2020 (https://www.loc.gov/preservation/digital/formats/fdd/fdd000318.shtml).

11. For DOCX text files: be guided by the description DOCX Transitional (Office Open XML), ISO 29500: 2008-2016, ECMA-376, Editions 1-5, published by the Library of Congress, updated 2017 (https://www.loc.gov/preservation/digital/formats/fdd/fdd000397.shtml).

12. For XML text files: be guided by the ANSI / NISO z39.96-2019 standard, JATS: Journal Article Tag Suite, Version 1.2, published by the National Information Standards Organization, approved in February 2019 (https://www.niso.org/publications/z3996-2019-jats).

13. For ARC related files: be guided by the description ARC_IA, Internet Archive ARC File Format, published by the Library of Congress, updated February 2008 (https://www.loc.gov/preservation/digital/formats/fdd/fdd000235.shtml).

14. For WARC related files: be guided by the WARC description, Web ARChive file format, published by the Library of Congress in February 2008 (https://www.loc.gov/preservation/digital/formats/fdd/fdd000236.shtml).

Useful tools and tips

This section lists some tools that may be of interest or useful to facilitate the processes of transformation and management of open formats:

1. FAIRassist (https://fairassist.org): has a collaboratively fed directory that includes questionnaires, guides, and automated validation tools that facilitate the implementation of FAIR principles in research data.
2. Top 10 FAIR Data & Software Things (https://librarycarpentry.org/Top-10-FAIR): made up of the list, grouped by subject areas, which contains various learning resources, data, relevant information, systems and links aimed at the scientific community, so that you can enter into the implementation of FAIR.
3. How to make your data FAIR (https://www.openaire.eu/how-to-make-your-data-fair): is a reference guide for researchers, published by the OpenAIRE organization, that describes steps and recommendations of interest, as well as a section of training materials for the adoption of the FAIR principles for research data.