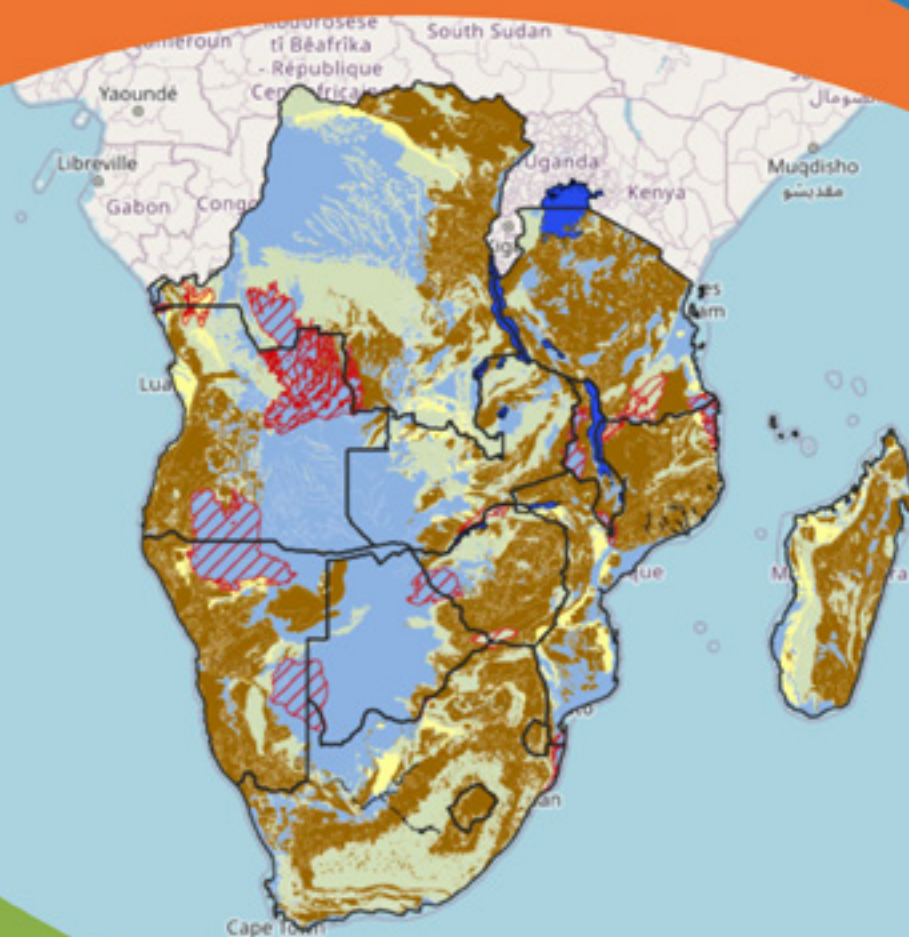




GROUNDWATER MANAGEMENT INSTITUTE

# GOOD PRACTICES FOR SHARING GROUNDWATER GEOSPATIAL DATA



**August 2020**

Authors: Hans van der Kwast (IHE-Delft), Arnaud Sterckx (IGRAC) and Raquel Sousa (IGRAC)

Acknowledgment: This document was developed as part of the project “**Expansion of the SADC Groundwater Information Portal (SADC-GIP)**”, commissioned by the SADC Groundwater Management Institute (SADC-GMI) and executed by the International Groundwater Resources Assessment Centre (IGRAC) in collaboration with IHE-Delft Institute for Water Education (IHE-Delft).



This document is published under Creative Commons license Attribution-ShareAlike 4.0 International (CC BY-SA 4.0). Under this license, credit must be given to the creator, and adaptations must be shared under the same terms. More information on the terms and conditions is available at <https://creativecommons.org/licenses/by-sa/4.0/>.

Preferred citation:

***SADC-GMI, IHE-DELFT and IGRAC (2020) Good practices for sharing groundwater geospatial data.***



# CONTENTS

1	INTRODUCTION	1
2	INTEROPERABILITY AND DATA FORMATS	2
2.1	Vector data files	2
2.2	Raster data files	3
2.3	Spatial databases	4
2.4	Style files	4
2.5	Project Files	4
3	METADATA	5
4	SDI	6
4.1	Interoperability	6
4.2	Metadata	6
4.3	Permissions and licenses	6
4.4	Data policy	7
5	REFERENCES	8

# 1

# INTRODUCTION

Groundwater is a vital resource for the SADC region. According to groundwater governance principles, it is important that all stakeholders receive appropriate information and access to groundwater data in order to participate in the sustainable management, development and protection of this resource<sup>1</sup>. However, regional capacity assessments carried out between 2017 and 2020 have highlighted that the sharing of groundwater data and information is often not effective<sup>2</sup>.

Ranging from borehole data, monitoring data and geophysical data to precipitation data, land use, and groundwater modelling data, groundwater data are for a very large part geographic data. It means that they have geographic coordinates (longitude, latitude and sometimes elevation) by which they can be referenced in a map. Also called geospatial data, georeferenced data or simply geodata, geographic data can be visualised, processed and analysed in Geographic Information Systems (GIS).

For this reason, it is important that groundwater professionals are properly trained and capacitated in using GIS. Besides on-site training, there are several resources online to learn the basics of GIS, become familiar or reach an advanced level with GIS software. A selection of online training material on QGIS, a mainstream GIS software is given in the box below.

The present document addresses the sharing of geospatial data between organisations and individuals engaged in groundwater management. It provides a set of good practices to both data providers and data users to ensure that data are shared properly. Getting these good practices into application would prevent common issues faced by groundwater specialists, such as: “I can’t open this datafile in my GIS software”, “These borehole data are not located at the right place in my map”, “Is this groundwater level expressed in depth below the surface or in elevation above the sea level?”

This document is structured in three sections. The first section covers interoperability issues between GIS programs and formats. The second section addresses the use of metadata, which are necessary for data users to know what data they are dealing with and how they can use it adequately. The third and last section presents how Spatial Data Infrastructures (SDI) can be implemented to facilitate the sharing of data within or among organisations.

---

<sup>1</sup> <http://www.groundwatergovernance.org>

<sup>2</sup> [http://sadc-gmi.org/wp-content/uploads/2019/07/State-of-GW-data-in-SADC\\_20190131.pdf](http://sadc-gmi.org/wp-content/uploads/2019/07/State-of-GW-data-in-SADC_20190131.pdf)

## **AN INDEX OF FREE QGIS MATERIAL**

<https://www.qgis.org/en/site/forusers/trainingmaterial/index.html>

## **QGIS FOR HYDROLOGICAL APPLICATIONS**

<http://www.gisopencourseware.org>

## **QGIS TUTORIALS AND TIPS**

<https://www.qgistutorials.com/en/>

## **VIDEO TUTORIALS**

<https://www.youtube.com/user/klakar70>

<http://www.youtube.com/c/hansvanderkwast>

Selection of online training material on QGIS

# 2

# INTEROPERABILITY AND DATA FORMATS

Most groundwater professionals with GIS experience are familiar with QGIS or ArcGIS, two desktop-based GIS applications with similar functionalities. While ArcGIS is a commercial software running with licenses, QGIS is open-source and can be downloaded and used for free. Apart from these two popular applications, many other applications are available<sup>3</sup>. It is important to consider that GIS also encompass web services, such as OpenStreetMap (OSM) or Google Maps, Google Earth, SDI (section 4), scripting languages, libraries, geospatial databases, web map servers, etc. All these different tools come with a very large amount of GIS data formats, which results in interoperability issues, as not all data formats are supported by all GIS applications. An extreme case is when a specific software is needed only to open the data, a situation known as vendor lock-in.

For desktop-based GIS applications, interoperability issues concern 5 types of files:

1. Vector data files
2. Raster data files
3. Style files
4. Spatial database files
5. Project files

Before opening a file, it is important to understand what it is and what it contains. Without metadata, files can be identified at their extension<sup>4</sup>.

## 2.1 VECTOR DATA FILES

Vector data are point, line or polygon features with associated attributes (Figure 2.1). They are one of the two primary types of geographic data in GIS, the other being raster data.



Figure 2.1 Examples of vector data layers.

The most popular format for vector data is probably the ESRI shapefile. Shapefiles consist in a minimum of 3 files, with different extensions:

- .shp: contains the feature geometry (point, line or polygon and associated georeferenced nodes)
- .shx: contains a positional index of the feature geom-

etry to allow seeking forwards and backwards quickly

- .dbf: contains the attribute table. It can be opened in a spreadsheet programme such as Excel.

It is impossible to use a shapefile without these 3 files. Other files can complete a shapefile. Two important ones are:

- .prj: contains the coordinates system and projection information in plain text.
- .cpg: used to identify the character encoding to be used. This file is needed to accommodate diacritics. For French for example the default UTF-8 should work, while for Portuguese ISO-8859-1 should be used.

If available, these two files should be shared along with the three mandatory files. It is important that all files have the same name (Figure 2.2). If not, GIS applications will not be able to recognize them as one shapefile. Since shapefiles are made of multiple files, it is advised to compress them into one file, such as .zip. Once compressed, shapefiles can be shared as one file, and they can be sent/downloaded more rapidly. Shapefiles need to be uncompressed before using. For compressing/uncompressing zip files, a free and open-source software like 7-Zip<sup>5</sup> can be used.

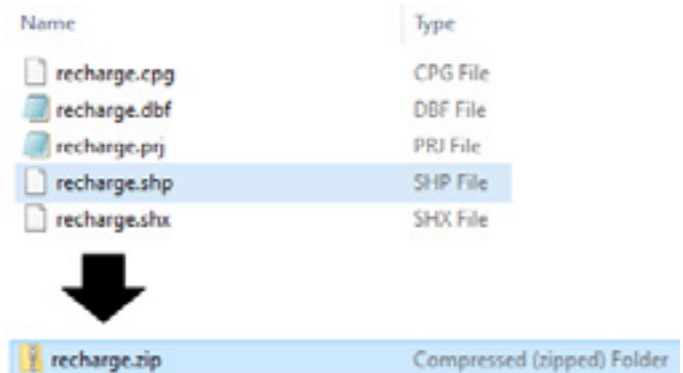


Figure 2.2 All files need to have the same name in order to be recognised as one shapefile. Compressing these files is a good practice for sharing.

<sup>3</sup> [https://en.wikipedia.org/wiki/List\\_of\\_geographic\\_information\\_systems\\_software](https://en.wikipedia.org/wiki/List_of_geographic_information_systems_software)

<sup>4</sup> File extensions are the 3 letters after the dot following file names. They give an indication on the nature or the purpose of files

<sup>5</sup> <https://www.7-zip.org>

Besides being made of multiple files, shapefiles have some limitations<sup>6</sup>, including the limitation of attribute names to 10 characters, which often results in inexplicit attribute names. There are around 100 alternative formats to shapefiles<sup>7</sup>. One is for instance the Keyhole Markup Language (KML). KML was originally developed as Google Earth<sup>8</sup> file format but is now read in several other GIS applications. KML can be converted from and to shapefile in most desktop-based GIS applications. One advantage of KML files is that they can be easily opened online in Google Earth, in case the user has no desktop-based GIS software.

Spreadsheets can also be used to store point data. Each column represents one attribute, while each row represents one point feature. Two columns need to store the coordinates of the points. Spreadsheets can be easily edited in Excel or free and open-source counterparts, like OpenOffice Calc. Some desktop-based GIS applications like QGIS allow the direct import of spreadsheets. In other cases, spreadsheets have to be exported in CSV format, which can be done in all spreadsheet programs. CSV is a human-readable format where columns are separated by comma (Figure 2.3).

LONG,	LAT,	ELEV,	DEPTH,	COMPLDATE,	COMPLTD_YR
31.72795,	-27.3533,	384,	55,	01/10/1987,	1987
31.37113,	-26.41950,	841,	42.7,	08/03/2004,	2004
31.3992,	-26.62,	382,	64.1,	28/01/1992,	1992
31.93514,	-27.27934,	140,	76.2,	07/09/1965,	1965
31.09099,	-26.26422,	1430,	-9999,	NO DATA,	-9999
31.38841,	-25.77611,	433,	84,	NO DATA,	-9999
31.47734,	-26.3254,	741,	61,	20/11/1990,	1990
31.57987,	-26.41925,	450,	120,	21/01/1989,	1989

Figure 2.3 Example of csv file.

The issue with spreadsheets is that they don't store the coordinated system. This has to be specified by the data provider in the header or in a separated metadata file (see section 3). If not specified, users might not be able to find out the coordinated system and to properly import the data in their GIS application. Like for other vector file formats, it is always recommended to overlay imported data with other data layers, to check if they are plotted at the right position, because projection issues are very frequent in GIS. Online maps such as from OpenStreetMap can be used as references.

CSV files have another downside: they don't always use comma as a separator. A different separator can be used, depending on the language settings of the computer. In languages where the comma is used as a decimal separator, the software will automatically choose another separator like semicolons. This is likely to be the case on computers set up in French or Portuguese. Other separators can also be assigned in case cells contain the original separator. For instance, a cell containing the following address: "SADC Groundwater Management Institute, Bloemfontein, South Africa" will be split in 3 different columns if the separator is

<sup>6</sup> <http://switchfromshapefile.org>  
<sup>7</sup> <https://gdal.org/drivers/vector/index.html>  
<sup>8</sup> <https://www.google.com/intl/en/earth/index.html>  
 (Google Earth is an interactive globe where you can overlay local KML files and data from the internet)  
<sup>9</sup> <https://gdal.org/drivers/raster/index.html>

a comma. A good practice is to always check the content of CSV files in a text editor like Notepad before importing them in a GIS software.

## 2.2 RASTER DATA FILES

Raster data are grids in which each cell has a given value (Figure 2.4). They are one of the two primary types of geographic data in GIS, the other being vector data.

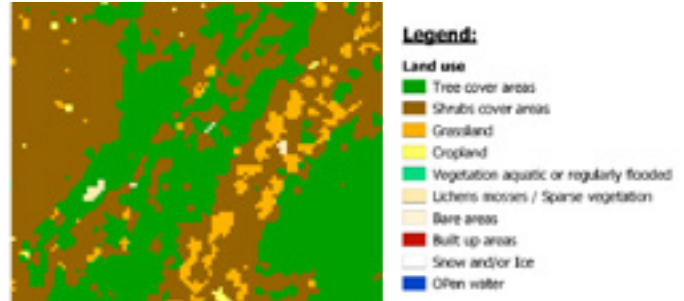


Figure 2.4 Example of a raster data layer.

As for raster data, there are several formats available<sup>9</sup>. The most commonly used are the GeoTIFF and the Arc/Info Binary or ASCII Grid. GeoTIFF is an extension of the TIFF (Tagged Image File Format) format, which is used to store raster graphic images. In addition to the image, the GeoTIFF format stores georeferencing information (e.g. map projection, coordinate systems, ellipsoids, datums) that allows the raster to be properly referenced in a GIS application. Sometimes the georeferencing information is stored in a separated file with the extension .tfw, also called World File. World Files must have the same name as the main TIFF file and must be stored and shared along. The Arc/Info Binary GRID is a proprietary format of ESRI. It can be read but not edited in non-ESRI applications. Like the ESRI shapefile, the Arc/Info Binary GRID is stored in several files contained in at least two directories: the name directory and an info directory. Arc/Info Binary GRID can be opened through the header file, recognisable at the .hdr extension. There is an open, editable alternative to the Arc/Info Binary GRID: the Arc/Info ASCII GRID. This human-readable format can be edited in a text editor, as illustrated in the box below.

ncols	4	• ncols and nrows are the numbers of columns and rows, respectively.
nrows	4	
xllcorner	0.0	
yllcorner	0.0	
cellsize	50.0	• xllcorner and yllcorner are the western (left) x-coordinates and southern (bottom) y-coordinates, such as easting and northing. When the data points are cell-centered xllcenter and yllcenter are used to indicate such registration.
NODATA_value	-9999	• cellsize is the length of one side of a square cell.
<pre>1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1</pre>		
• nodata_value is the value that is regarded as "missing" or "not applicable"; this line is optional, but highly recommended as some programs expect this line to be declared.		
• The remainder of the file lists the raster values for each cell, starting at the upper-left corner and delimited using a single space.		

Example of Arc/Info ASCII GRID file format

## 2.3 SPATIAL DATABASES

Dealing with multiple vector and raster files can be tedious and confusing. A better solution is to store geospatial data in a spatial database. A spatial database is optimized to store and query several data. Spatial databases can be of two types: personal or multi-user.

A personal database is designed for individual use. It allows storing all data layers in one database file, which is very easy to share. Some databases even allow to save the project file. Spatial databases are supported by desktop-based GIS applications and don't require any additional software to run. Here are some examples of geospatial databases for storing vector and raster files that can be opened in ArcGIS and QGIS:

- ESRI File Geodatabase (.gdb) consists in several files stored in a folder. It has a 1TB file limitation. It replaced the ESRI Personal Geodatabase (.mdb).
- Spatialite (.SL3, .SQLITE) is an open format that extends the SQLite database with Spatial SQL capabilities.
- GeoPackage (.GPKG) is an open, standards-based, platform-independent, portable, self-describing, compact format for transferring geospatial information. Published as an OGC<sup>10</sup> standard, it is supported by many applications. In QGIS, even project files can be stored in a GeoPackage (Figure 2.5), which makes it very easy to share via file transfers, USB stick or e-mail.

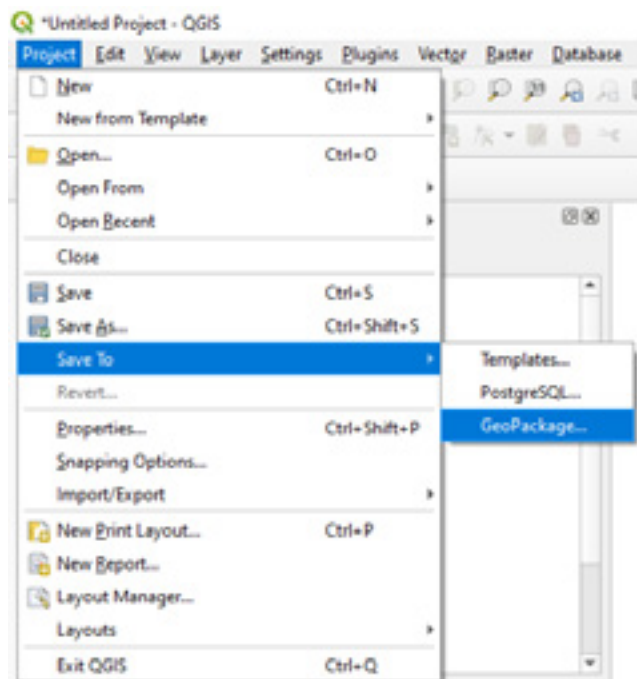


Figure 2.5 Saving a QGIS project in a GeoPackage

Multi-user spatial databases are meant to be used by several users within one organisation or across multiple organisations. One example is PostGIS, an extension of the object-relational database PostgreSQL. PostGIS is open-source and is used by many other GIS applications. Multi-user spatial databases are at the core of Spatial Data Infrastructures (section 4).

## 2.4 STYLE FILES

Style files describe how vector and raster data display in a GIS application. When sharing geospatial data, it is important to share the associated style file to allow other users to display the data in the same way (Figure 2.6).

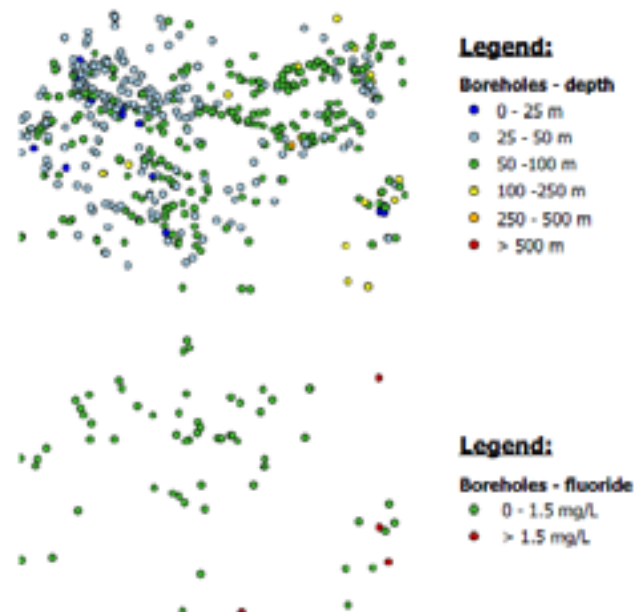


Figure 2.6 Two different legends have been applied to the same borehole data layer, highlighting the depth (above) and the fluoride content (below) in the boreholes.

Common styles file are for instance .lyr files and .qml files, which can be opened in ArcGIS and QGIS respectively. Unfortunately, .lyr files are not supported in QGIS and .qml files are not supported in ArcGIS. It is recommended to save styles in .sld files (for Style Layer Descriptor), an OGC standard supported by QGIS, ArcGIS and many other GIS applications.

## 2.5 PROJECT FILES

Project file formats are software-specific, they don't aim at interoperability. ArcGIS project files like .mxd files can't be opened in QGIS and QGIS project files like .qgs and .qgz can't be opened in ArcGIS. Another issue is that project files don't store the layers, but rather the location of the layers on the disk (path). For these reasons, it is not advised to share project files.

<sup>10</sup> <https://www.opengeospatial.org/docs>

# 3

# METADATA

Metadata are data about the data. They provide information about the content, quality, structure, accessibility and other characteristics of the data (GSDI, 2013). Detailed and robust metadata must document the project, datasets, and services in such a way that data can be transmitted, interpreted, reused, and understood. Without metadata, it won't be possible for others to use the data properly. For example, it won't be possible for someone to locate boreholes stored in a spreadsheet without being given the coordinate system. A groundwater level map will have little value if the original groundwater level data and the calculation method are not specified. It might not be clear for everyone that a shapefile attribute column named "ELEV\_M\_DEM" gives the elevation from a Digital Elevation Model in meters, and nobody will be able to guess which DEM was used, if not mentioned in the metadata.

Metadata can include:

- the coordinate system
- the original data used to produce the layer
- the method used to produce the layer
- the measurement units
- the resolution
- the date
- the author(s)
- the owner(s) of the data
- the license or restrictions to share and use the data

In case of shapefiles, it is also necessary to provide the full names of attributes, which are limited to 10 characters.

COMPLTD_YR	SWL_M	SWL_DATE	YIELD_L/S
------------	-------	----------	-----------

Figure 3.1 Example of attribute names in a shapefile

These metadata can sometimes be embedded in the data file itself. In QGIS, the metadata of raster and vector layers can be edited (Figure 3.2).

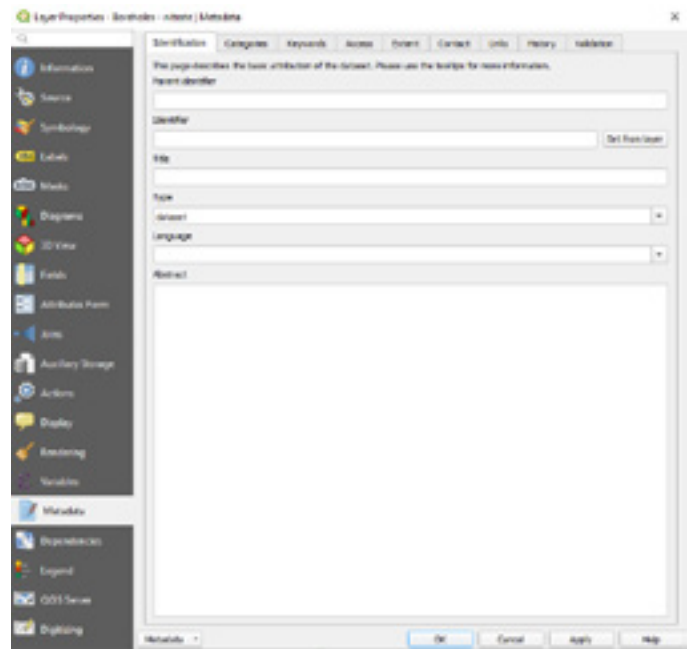


Figure 3.2 Metadata editor in QGIS

In other cases, like data stored in spreadsheets, metadata have to be written in a separated file and shared along. Metadata is even more important when data is shared outside an organisation or a country, as conventions and practices will likely differ. For the sharing of data at this level, the use of Spatial Data Infrastructure (SDI) is recommended, as they rely on extensive metadata standards.



# 4

# SPATIAL DATA INFRASTRUCTURE

A Spatial Data Infrastructure (SDI) is an online environment supporting the sharing of geospatial data, where users can:

- Discover data through a search engine and a catalogue
- Visualise data through interactive maps or infographics
- Evaluate data before using it, through metadata
- Access data through web-based services, mobile apps or GIS applications

Just like a road network supports the transport of people and goods, a SDI supports users to share geospatial data. Like with a road, there are rules to make sure that everyone uses it properly, and there is flexibility for a wide range of users to use it their way. Examples of SDI in the SADC region include the SADC Groundwater Information Portal<sup>11</sup>, the ORASECOM GIS server<sup>12</sup> or the Malawi Spatial Data Portal<sup>13</sup>.

## 4.1 INTEROPERABILITY

An SDI relies on a multi-user geospatial database and OGC standards and services, which solves most of the interoperability issues explained above. In an SDI like the SADC Groundwater Information Portal, registered users can share the following data:

- Vector data, as shapefiles
- Raster data, as GeoTIFF
- Other data formats, such as spreadsheets, images, PowerPoint presentations, videos and pdf or word documents

Style files have to be uploaded in sld. format. Documents can be associated with data layers, which is useful for example for linking a dataset with a corresponding methodology report.

## 4.2 METADATA

An SDI also integrates the sharing of extensive metadata, in a standardized way. In the SADC-GIP, several metadata fields can be filled in. These metadata are necessary for other users to find the data in the catalogue and to know how to use it properly.



Figure 4.1 Metadata in the SADC groundwater information portal (SADC-GIP)

## 4.3 PERMISSIONS AND LICENSES

An SDI allows users to control the sharing of their data through permissions and licenses. In the SADC-GIP, any data provider decides:

- Who can view the data
- Who can download the data
- Who can change metadata
- Who can edit the data
- Who can edit styles
- Who can manage the data (update, delete, change permissions, publish/unpublish)

In the metadata, data providers can also add a license on the data. This also applies to open data, which are defined as 'data that can be freely used, reused and redistributed by anyone - subject only, at most, to the requirement to attribute and share alike'<sup>14</sup>. With a license, the data provider specifies how the data can be used by others. Some commonly used Creative Commons licenses<sup>15</sup> are reported in the box below.




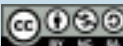

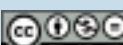
<sup>11</sup> <https://sadc-gip.org>

<sup>12</sup> <http://gis.orasecom.org>

<sup>13</sup> <http://www.masdap.mw>

<sup>14</sup> [http://opendatahandbook.org/guide/nl\\_BE/what-is-open-data](http://opendatahandbook.org/guide/nl_BE/what-is-open-data)

<sup>15</sup> <https://creativecommons.org>

SYMBOL	DESCRIPTION
	<a href="#">CC BY</a> - This license allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator. The license allows for commercial use.
	<a href="#">CC BY-SA</a> - This license allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator. The license allows for commercial use. If you remix, adapt, or build upon the material, you must license the modified material under identical terms.
	<a href="#">CC BY-NC</a> - This license allows reusers to distribute, remix, adapt, and build upon the material in any medium or format for non-commercial purposes only, and only so long as attribution is given to the creator.
	<a href="#">CC BY-NC-SA</a> - This license allows reusers to distribute, remix, adapt, and build upon the material in any medium or format for noncommercial purposes only, and only so long as attribution is given to the creator. If you remix, adapt, or build upon the material, you must license the modified material under identical terms.
	<a href="#">CC BY-NC-ND</a> - This license allows reusers to copy and distribute the material in any medium or format in unadapted form only, and only so long as attribution is given to the creator. The license allows for commercial use.
	<a href="#">CC BY-NC-ND</a> - This license allows reusers to copy and distribute the material in any medium or format in unadapted form only, for noncommercial purposes only, and only so long as attribution is given to the creator.

### Overview of Creative Commons licenses

For proprietary data, a copyright statement can be used.

It is recommended that the sharing of groundwater data within or among organisations be guided by a data governance policy, which is “a set of guidelines for ensuring that an organization’s data and information assets are managed consistently and used properly. Such guidelines typically include individual policies for data quality, access, security, privacy and usage, as well as roles and responsibilities for implementing those policies and monitoring compliance with them”<sup>16</sup>. Whereas most organisations in SADC have a policy on information sharing, few have a policy on data sharing. The FAIR principles offer a good starting point to develop data sharing policies. FAIR is the acronym of Findable, Accessible, Interoperable and Reusable. The proper use of file formats and standards, metadata and SDI, as explained above, supports the fulfilment of these 4 principles. The FAIR principles also invite organisations to share data as open data. In SADC, some groundwater data are still considered as sensitive and are therefore not shared publicly. It can be argued that these groundwater data are collected by public institutions funded with public money, and taxpayers should be given access to these data, as long as they do not infringe the privacy and safety of citizens. Open data also provide more transparency and a better knowledge of groundwater resources to citizens, researchers or private companies, enabling these stakeholders to participate in the sustainable management of groundwater in an informed manner. Furthermore, Open Data authorize the reuse of data, allowing the creation of more knowledge and products.

According to the definitions of FAIR and Open Data, it is important to note that maps in pdf or other non-editable formats are not open, because the original data can’t be reused. For example, a pdf map showing the location of boreholes doesn’t allow the reuse of borehole locations and other borehole data for further analysis and applications.

Eventually, a data governance policy should outline:

- Who can access the data (if not Open Data)
- Under which conditions (cfr licenses)
- How the data are shared (e.g. through an SDI, by email, in which format)

<sup>16</sup> <https://searchcompliance.techtarget.com/definition/Data-governance-policy>

# 5

# REFERENCES

- **Creative Commons:** <https://creativecommons.org/>
- **Geolexia**, 2020. The authoritative glossary for geographic information technology from ISO/TC 211. <https://isotc211.geolexia.org/concepts/202/#entry-lang-eng>
- **GDAL/OGR:** <http://www.gdal.org>
- **GEOSS:** <http://www.epa.gov/geoss/>
- **GSDI**, 2013. The SDI Cookbook, Global Spatial Data Infrastructures Association (GSDI), [http://www.gsdi-docs.org/GSDIWiki/index.php/Main\\_Page](http://www.gsdi-docs.org/GSDIWiki/index.php/Main_Page)
- **INSPIRE:** <http://inspire.jrc.ec.europa.eu/index.cfm>
- **OGC**, 2013. Open Geospatial Consortium, <http://www.opengeospatial.org>
- **UNECA**, 2003. The SDI handbook for Africa, United Nations Economic Commission for Africa (UNECA), <http://repository.uneca.org/handle/10855/5097>
- **UNSDI:** <http://www.ungis.org/content/united-nations-spatial-data-infrastructure-unsdi>
- **USGS**, 2013. National Climate Change & Wildlife Science Centers Data Sharing Policy, <https://nccwsc.usgs.gov/content/data-policies-and-guidance>





## **SADC GROUNDWATER MANAGEMENT INSTITUTE**

Institute for Groundwater Studies  
Building Dean Street,  
University of the Free State  
205 Nelson Mandela Drive  
Bloemfontein, South Africa

**[www.sadc-gmi.org](http://www.sadc-gmi.org)**  
**[info@sadc-gmi.org](mailto:info@sadc-gmi.org)**  
**(+27) 51 401 7734**