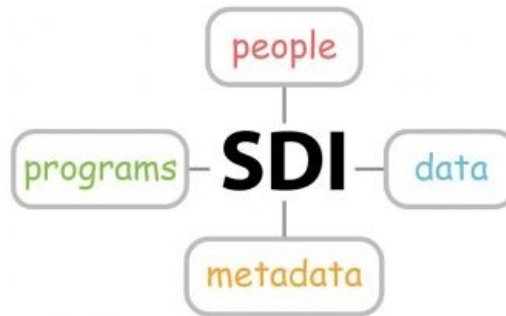


# Building an SDI with FOSS



A spatial data infrastructure (SDI) seeks to harmonise users with the provision of data, metadata, software and computational services to provide a productive and flexible environment for working with geospatial data. In an enterprise environment, the development of an SDI constitutes the foundations of the GIS department or group in your organisation. The Open Geospatial

Consortium (OGC) provides a number of standards for various elements of an SDI. In this article, we will review the software options provided by the FOSSGIS (Free and Open Source software for Geographical Information Systems) community for developing an OGC standards compliant spatial data infrastructure (SDI). This will provide the foundation for future articles where we will delve into some of the individual projects described below.



Let us start our grand FOSSGIS SDI tour with the basics: The **spatial data repository**. In the previous edition of GIS Pro, we took a look at PostGIS and PostgreSQL via an interview with PostGIS co-founder, Paul Ramsey. For storage of vector data in the enterprise, PostGIS is an excellent choice. It runs on all major platforms (Windows, Linux, MacOS) and provides an enterprise-ready data store for vector data, capable of providing

for multiple concurrent users and hosting these larger datasets. PostGIS supports representation of features in both the OGC Simple Feature Specification or as true topology. It also supports representing true curves. Although PostGIS is capable of storing raster data too, it is more usual to use a traditional file system based approach for hosting raster data. For **file based storage**, the venerable GDAL library deserves a mention as its command line tools allow for various transformation services of raster data, including enabling one to generate hierarchical tile mosaics (TMS) stores from source raster data.

Once you have established a data store, the next consideration is the annotation and publication of this data with standards compliant metadata in a **metadata catalogue**. This is referred to as the catalogue service of the SDI. Typically an SDI should publish data using ISO based institutional standards such as INSPIRE, ISO 19115 etc. made available by the Catalogue Services for the Web (CSW) standard. The flagship product here is GeoNetwork, which provides tools to manage metadata for geospatial data sources, search that metadata and browse the related datasets using a web map viewer. PyCSW (which, along with GeoNetwork, provides a reference implementation for the OGC CSW standard) is a library for publishing and managing CSW-compliant metadata. Platforms such as GeoNode and CKAN provide spatial data storage services, with PyCSW providing catalogue services, to form a spatial content management system, or an SDI in a box.

There is a need to direct users to the data itself after it has been discovered in the catalogue. This functionality is provided by the **spatial data service** component of an SDI. GeoServer, Mapserver and QGIS Server are great examples of FOSSGIS data publishing engines. These spatial data servers make data stored in the spatial data repository available via a number of standards based protocols: The web feature service (WFS) for serving vector feature data, web coverage service (WCS) for serving raster coverages and the web mapping service for publishing ready-to-consume cartographic renderings of one or more datasets.

Users within an SDI environment will often need to transform and manipulate data between data formats, projections and map synonymous attributes between different data sources when merging or importing datasets. In the FOSSGIS world, Proj4 is a key project for the provision of coordinate reference system **transformation services** and is used by many popular GIS tools such as GDAL and QGIS. GDAL (Generic Data Access Library) provides the capability for reading and writing a huge variety of raster and vector (via its OGR sub-system) data formats. It is also used by various commercial GIS applications.

Many SDI implementers will also look to deploy **web-based processing services** via the Web Processing Service (WPS) standard. A WPS can offload resource intensive tasks to a server and away from users' desktops. Additionally it offers a way to have processing take place 'near the data' instead of requiring that users retrieve large datasets locally to do analysis. The FOSSGIS community provides some excellent options for WPS: Zoo and PyWPS being notable examples.

With the above elements in place, a key remaining element is to leverage the SDI to provide GIS services to users via desktop and web-based **GIS applications**.

For desktop users, there are a number of mature, feature rich GIS applications. GRASS GIS is one of the oldest GIS projects, commercial or otherwise (over 30 years). GRASS is still under active development and its developers have ensured that the application remains relevant with its rich assortment of analytical tools, topological data model and host of other features. gvSIG and QGIS, although newer projects than GRASS, are both mature applications and also provide mobile applications as sub-projects. They offer a more modern interface than GRASS. QGIS and the Community Edition of gvSIG can interface with GRASS to leverage its analytical capabilities. There are a number of other FOSSGIS desktop GIS applications worth exploring: uDIG and Orfeo Toolbox (OTB) being good examples. uDIG specialises in being a great OGC web services client. OTB is an extremely powerful remote sensing application. The nature of these

projects being open source software means that your desktop GIS choice need not be an all-or-nothing affair. Projects like QGIS also act as a front end to OTB, GRASS and SAGA, giving you a 'best of all worlds' approach.

When users need to discover and **access the SDI via a web browser**, the FOSSGIS community has developed a number of useful applications. Probably most well known are the javascript libraries OpenLayers and Leaflet. The former provides an extremely versatile set of functionality for making richly functional web applications. Leaflet, on the other hand, aspires to be a very lightweight Javascript GIS browser that can be extended via plugins as needed. These are both build-your-own toolkits, though there are a number of ready to run web applications to serve many different use cases.

In almost all cases, **OGC Web Services** are the underpinnings of these applications and the data exchange between applications, so if you have not already done so, it is worth familiarising yourself with the various OGC standards if you intend to build out an SDI. FOSSGIS developers tend to eschew proprietary formats and protocols as they are hard to interoperate with - an SDI should facilitate data exchange between the different components as easily as possible, while also making it easy to open services to the outside world. Open standards (such as those published by OGC) go a long way to facilitate this.

So, given the whirlwind tour above, you may be wondering "what is missing"? There are still holes in the matrix of tools and applications provided by **OSGeo** (the umbrella project for most of the above-mentioned applications). These holes are most noticeable when you start entering into vertical markets. The more specialised the domain (e.g. mining, civil engineering, natural resource management), the less likely you are to find ready to use schemas, applications and data to incorporate into your SDI. That said it is always worth doing some research before discarding the idea of using a FOSSGIS based stack - you may well find something that fits your needs well. In cases where vertical markets are not well accommodated, many users 'build their own' applications using the numerous FOSSGIS tools, and it is not uncommon to see a number of organisations in a given sector banding together to fund the development of open source tools to support their needs.

If you are interested in finding out more about the options available to you, it is worth attending the annual FOSSGIS user conferences - this year it will be held at the end of August in Dar es Salaam (<http://2018.foss4g.org>), or take a virtual tour of the projects at <http://osgeo.org>.

*This article was published in GIS Professional June 2018*

### **About the Author**

*Tim Sutton is QGIS project chair and director at Kartoza Pty Ltd. This article represents his personal views and opinions and not those of his employer (Kartoza Pty Ltd.) or the QGIS Project (<http://qgis.org>).*