FitsMap: A Simple, Lightweight Tool For Displaying Interactive Astronomical Image and Catalog Data

Ryan Hausen\textsuperscript{a,}\textsuperscript{*}, Brant E. Robertson\textsuperscript{b}

\textsuperscript{a}Department of Computer Science and Engineering, University of California, Santa Cruz, 1156 High Street, Santa Cruz, 95064, CA, USA
\textsuperscript{b}Department of Astronomy and Astrophysics, University of California, Santa Cruz, 1156 High Street, Santa Cruz, 95064, CA, USA

Abstract

The visual inspection of image and catalog data continues to be a valuable aspect of astronomical data analysis. As the scale of astronomical image and catalog data continues to grow, visualizing the data becomes increasingly difficult. In this work, we introduce \textit{FitsMap}, a simple, lightweight tool for visualizing astronomical image and catalog data. \textit{FitsMap} only requires a simple web server and can scale to over gigapixel images with tens of millions of sources. Further, the web-based visualizations can be viewed performantly on mobile devices. \textit{FitsMap} is implemented in Python and is open source (https://github.com/ryanhausen/fitsmap).

Keywords: Astronomy web services, Astronomy data visualization, Astronomy data analysis, Human-centered computing Scientific visualization, Human-centered computing Visualization toolkits

1. Introduction

Astronomical image data is inherently visual, and visual inspection and interpretation remain vital tools in the scientific process in astronomy. Upcoming telescopes like the James Webb Space Telescope [JWST; for a review, see 31], Nancy Grace Roman Space Telescope [34, 2], and Vera Rubin Observatory [20, 19] will produce larger and deeper images of space than ever before. Specialized tools for visualizing and interacting with large scale astronomical images will enable a more rapid transition from observation to analysis for these facilities.

Visualizing large astronomical images is not a new problem, and several tools have been developed to meet this need. There are desktop tools like \textsc{DS9} [21] and \textsc{Aladin} [10], which are \textit{thick} clients that render local images or fetch images from a remote server. The term \textit{thick client} refers to programs that have little to no reliance on a remote server to perform the computational tasks associated with the software. Advances in web technologies have also enabled the development of \textit{thin} clients for astronomical image data visualization. \textit{Thin} clients are dependent on a separate server to perform their computational tasks. An advantage to using thin clients is that by offloading heavier computational tasks to a remote server the client hardware requirements are far less powerful than the server requirements and can even include mobile devices. Some examples of thin clients developed for astronomical image and catalog visualization include \textsc{VisiOmatic} [8, 7], \textsc{Aladin Lite} [9], \textsc{RCSed} [11, 23], the \textsc{GAIA} archive visualization service [28], \textsc{Toyz} [29], \textsc{WorldWide Telescope} [32], \textsc{ASTRODEEP} [13, 37], and \textsc{ESASky} [16]. More recently, the popularity of the Python [33] programming language has inspired a series of new \textit{thick} and \textit{thin} clients that leverage Python and the Jupyter Lab/Notebook (iPython; Khuver2016) ecosystem. Some examples include \textsc{Vizic} [40], \textsc{LSSGalPy} [4], \textsc{Astro Data Lab} [22], \textsc{Jovial} [3], and \textsc{Jdaviz} [30].

This work introduces a new tool called \textit{FitsMap}. \textit{FitsMap} is a thin client that improves on previous methods by only requiring a simple web server for offloading computational tasks. \textit{FitsMap} is designed to work for a simple use case where the user has images (\textsc{FITS} [38], \textsc{PNG} [36], or \textsc{JPEG} [35]), catalogs associated with the image data, and would like to view and possibly share that data interactively. A user can generate a website that displays image and catalog information using a single \textit{FitsMap} function. \textit{FitsMap} has already been deployed to visualize large simulated JWST [39] and Roman [15] datasets and machine-learning galaxy morphological analyses of large Hubble Space Telescope surveys [15].

The remainder of this paper is structured as follows. Section 2 describes the details of the design and methods of \textit{FitsMap}. Section 3 describes the performance of \textit{FitsMap} and, Section 4 reviews the contributions and future directions of \textit{FitsMap}.

2. Methods

The design philosophy behind \textit{FitsMap} is to render visualizing image and catalog data simple by minimizing the number of steps and technical knowledge required to go
from image and catalog files to an interactive display. The only requirements to generate and view the output from FitsMap are FitsMap itself and a web server. If viewing the output locally, FitsMap can also run the web server. If FitsMap is used to view image data alone a web server is not required, as the user can simply open the output index.html file using a browser and the processed images will render.

Catalog data present a unique challenge in the effort to minimize server requirements. Small catalogs can be transferred from the server to be clustered and rendered on the client, but large catalogs practically cannot. Large catalogs may either take too long to send usefully or exceed the memory capabilities of the client. Managing the catalog data on the server-side requires a more sophisticated web server that clusters and serves catalog data as needed. Leveraging the continuing decline in the cost of storage [27], FitsMap separates the catalog data into a set of tiles at different zoom levels. FitsMap then precomputes clusters of the catalog data at every zoom level for every tile (Section 2.1.2). Precomputing the clustering and tiling the data significantly reduce the computational requirements of both the server and client by never requiring the entire catalog to be loaded into memory.

The FitsMap architecture consists of two main components: the Map Generator and the Map Viewer. The Map Generator, implemented in Python, builds the website, including the Map Viewer files. The Map Viewer, implemented in Javascript, renders the webpage and fetches the image and catalog data.

2.1. Map Generator

The Map Generator parses a list of input image and catalog file locations and converts them into a format that the Map Viewer can render. The output of the Map Generator is a directory that contains the processed input image and catalog data and the supporting HyperText Markup Language (HTML), images, Cascading Style Sheets (CSS), and JavaScript. The Map Generator can process multiple files in parallel and can further parallelize the processing of each file. For example, image and catalog tiles can be generated in parallel. See Figure 1 for a graphical representation of the Map Generator. The Map Generator performs two major tasks when generating an interactive website: parsing image files and parsing catalog files. FitsMap dynamically generates the Map Viewer files using the image and catalog data information.

2.1.1. Parsing Image Data

The first step in parsing the image data is to convert the image data into a numpy array. The Map Generator converts FITS files using astropy [6] and converts traditional image files (PNG, JPG, etc.) using pillow [12]. The Map Generator builds tiled representations, in parallel if desired, of the array data at each zoom level in a structured set of directories compatible with the Map Viewer.

2.1.2. Parsing Catalog Data

The Map Generator can parse plain text files that are delimited (CSV, TSV, etc.) and where the first line contains the column names. Catalog source locations can be stored in either \([x, y]\) (image pixel coordinates) or \([\alpha, \delta]\) (right ascension and declination). If the coordinates are given in \([\alpha, \delta]\), then a FITS file containing a reference World Coordinate System (WCS) to translate the sky coordinates to image coordinates must be provided. After reading the catalog data, the Map Generator clusters the data at every zoom level using a python port [20] of the supercluster [25] JavaScript library. Then the Map Generator builds a tiled representation of the catalog source and cluster locations at every zoom level. The tiled catalog source locations are encoded and stored using the MapBox Vector Tile Format [21], a compact binary representation of structured data used to represent geometric objects in map tiles. Other source attributes (effective radius, magnitude, etc.) are stored in a separate JavaScript Object Notation (JSON) file for each source in a dedicated directory for each catalog.

2.2. Map Viewer

The Map Viewer consists of the supporting HTML, CSS, image, and JavaScript files that render the image and catalog data. The Map Viewer is built using Leaflet [1]. Leaflet is an open-source and flexible framework for displaying large amounts of data. Leaflet supports rendering image tiles, and to support tiled marker data FitsMap uses a custom-implemented Leaflet layer called L.GridLayer.TiledMarkers, which is packaged with FitsMap. The Map Viewer fetches and renders markers (sources and clusters) using the data stored in each of the encoded tile files; see Figure 2 for an illustration of image and catalog tiling. When a source marker is selected, the Map Viewer fetches additional information about the source from its corresponding JSON file (see Section 2.1.2 and Figure 1) and renders it in a pop-up above the marker (see Figure 3). The marker pop-ups support HTML to enable the inclusion images and styling using a custom CSS. The Map Viewer also includes a search function that can be used to query any of the parsed catalogs by their id column. The search functionality is implemented using a custom search backend to the popular Leaflet Control Search plugin [13] that is packaged with FitsMap. To search by id, the Map Viewer sends a GET request for a JSON file named \{id\}.json in each catalog’s extra JSON source file directory in catalog/assets (see Figure 2). If the request returns 404 (file not found), then the id does not exist. If the request returns 200 (OK), then the desired object does exist and the Map Viewer can extract location information stored in the file to pan and highlight the source.

3. Performance

As described in Section 2 to reduce the computational requirements of both the server and the client FitsMap
tiles the marker data such that the client only renders the markers that are currently in view and unclustered at the current zoom level. In this section, the performance of FitsMap is evaluated in two ways. First, the Map Generator is evaluated with respect to computational and storage requirements. Second, the performance of the Map Viewer is evaluated on both desktop and simulated mobile environments using website profiling tools.

3.1. Map Generator Performance

The Map Generator bears the largest computational burden within the FitsMap architecture. FitsMap tiles an image by taking subsets of the input image file and interpolates them into tile-sized images (256² pixels). FitsMap optimizes the tiling process in two ways. First, image tiling is parallelized over tiles to provide a significant speedup. Second, to ensure the entire input image does not have to load into memory FitsMap offers the option to limit how far out the map may be zoomed. FitsMap processes catalog data in three steps: parsing, clustering, and tiling (see Figure 1). Both parsing and tiling are parallelizable in FitsMap. The storage requirements of clustering and tiling source catalog information are not easily predicted before generating the map, but the storage requirements for the JSON storage of catalog sources with r rows and c columns can be estimated per source using

\[ M_j(r, c) = s(r) + s(r_{xy}) + s(c) + n_f + 6n_c + 30 \]  

(1)

where \( s(\cdot) \) is a function that returns the number of bytes required to encode the argument as a string (excluding quotation marks, i.e. \( s(3) = \text{sizeof}("3") = 1 \)), \( r_{xy} \) is the location of the source represented by row \( r \) in the image,
For example, a tile containing a single source and a single
source and $n$ the storage requirements per tile containing
to account for data type and repeating values. However,
ally, the Mapbox Vector Tile format is highly optimized
in a catalog is difficult a priori to store the source in
JSON. 24, totaling 55 + 30 = 85. This indicates that the cost to
Each term dependent on $r$ and $c$ in Equation 1 yields the
following: $s(r) = 8$, $s(r_{xy}) = 4$, $s(c) = 12$, $n_f = 7$, $6n_c = 24$, totaling $55 + 30 = 85$. This indicates that the cost to
store the source in JSON for FitsMap is 85 bytes.
Calculating the total storage cost of tiling and clustering a catalog is difficult a priori as clustering depends on the spatial distribution of sources in the image. Additionally, the Mapbox Vector Tile format is highly optimized to account for data type and repeating values. However, the storage requirements per tile containing $n_S$ individual sources and $n_C$ clusters of sources can be estimated using

$$M_t(n_S, n_C) \approx 99 + 53n_S + 56n_C. \quad (2)$$

For example, a tile containing a single source and a single

![Image Data](image1.png)

![Zoom Level](image2.png)

![Image & Catalog Data](image3.png)

Figure 2: An illustration of tiling image and catalog data at three zoom levels (0, 1, 2). (left) Illustration of tiling image data only. (right) Illustration of tiling image and catalog data. For simplicity, not all sources in the image are labeled with a marker. FitsMap computes the cluster assignments at every zoom level and stores them as markers in their respective tile. When the Map Viewer requests the catalog data associated with a particular tile, it will retrieve either a source marker or a cluster marker and renders it appropriately. Source image credit: NASA.

Figure 3: An example of the FitsMap interface. (a) The search function button that searches catalogs by id (see Section 2.2). (b) A marker pop-up displaying catalog data associated with the indicated source. (c) The Leaflet layer control that allows users to switch between display images and catalog overlays. Image Credit: [15].

3.2. Map Viewer Performance

The Map Viewer is written entirely in JavaScript and
only requires a web server to render the output from the
Map Generator. Map Viewer performance is measured
by rendering the website generated in Section 3.1 which contains a 50,000 pixel FITS image (10GB) and a sample catalog (176MB) containing 378,082 sources with 57 columns. Among five independent trials, generating the map on an Intel Xeon E5-2698 system with 232GB of RAM and processing the catalog and image in parallel with each process using six workers took 213.3522 ± 4.5727 seconds. The total data size of the input was approximately 10.2GB, and the total of the generated website was approximately 1.8GB. The main reason for the storage reduction is the conversion of FITS to PNG for the image tiles. The catalog data size grew from the input 176MB to approximately 1.5GB upon output, and constituted most of the output data.
2.9 seconds. Additionally, the Chrome performance profiling tool was used to monitor memory consumption during a session consisting of loading the page, displaying the catalog, and zooming in until clustering is no longer applied and all local source locations are rendered. In the trial run for both mobile and desktop, the Map Viewer never exceeded 16MB and stayed well within the RAM capacity of desktops and mobile devices.

4. Conclusion and Future Work

This work introduced FitsMap, a simple, lightweight tool for generating interactive web-based visualizations of astronomical images and their associated catalog data. FitsMap uses a novel approach to serve catalog data that precomputes cluster information for the entire image at all zoom levels and displays the cluster and source information based on currently visible image tiles. Advantages of precomputing and tiling the catalog data include 1) the visualization only requires a simple web server and 2) the produced website loads quickly and proves responsive for large catalogs, even on mobile devices. The techniques developed in this work have already been used on public-facing websites. Both [15] and [16] leverage FitsMap to generate interactive visualizations displaying their data products (See https://purl.org/fitsmap/morpheus and https://purl.org/fitsmap/dream respectively). Potential drawbacks to FitsMap include the increased storage requirements above the input image and catalog data sizes that owe to the tiling method. The FitsMap clustering and tiling requires preprocessing of the catalog data, but the preprocessing only needs to be performed once per website. The development of FitsMap is active and on-going. Future work includes allowing for more complex shapes in catalog data such as polygons, allowing users to change FITS image scaling on the fly in the Map Viewer, and minimizing storage usage.

5. Acknowledgements

We thank Daniel Eisenstein for suggesting the concept that eventually became FitsMap. The authors acknowledge use of the lux supercomputer at UC Santa Cruz, funded by NSF MRI grant AST 1828315, and support from NASA grant 80NSSC18K0563 and the NIRCam science team.

References

[1] Agafonkin, V., Cloudmade, 2022. Leaflet. URL: https://github.com/Leaflet/Leaflet
[2] Aleson, R., Armus, L., Bachelet, E., Bailey, V., Bartusek, L., Bellini, A., Benford, D., Bennett, D., Bhattacharya, A., Bohlin, R., Boyer, M., Bozza, V., Bryden, G., Calchi Novati, S., Carpenter, K., Casertano, S., Choi, A., Content, D., Dayal, P., Dressler, A., Doré, O., Fall, S.M., Fan, X., Fang, X., Filippenko, A., Finkelstein, S., Foley, R., Furlanetto, S., Kailari, J., Gaudi, B.S., Gilbert, K., Girard, J., Grady, K., Greene, J., Guhathakurta, P., Heinrich, C., Hemmati, S., Henden, D., Henderson, C., Henning, T., Hirata, C., Ho, S., Huff, E., Hutter, A., Jansen, R., Jha, S., Johnson, S., Jones, D., Kasdin, J., Kelly, P., Kinshan, R., Koekemoer, A., Kruk, J., Lewis, N., Macintosh, B., Madau, P., Malhotra, S., Mandela, K., Massara, E., Masters, D., McEnery, J., McQuinn, K., Melchior, P., Melton, M., Mennesson, B., Peeples, M., Penny, M., Perlmutter, S., Pisani, A., Plazas, A., Poleski, R., Postman, M., Ranc, C., Rauscher, B., Rest, A., Roberge, A., Robertson, B., Rodney, S., Rhoads, J., Rhodes, J., Ryan, Russell, J., Sahu, K., Sand, D., Scolnic, D., Seth, A., Shavaliwals, Y., Stierle, K., Smith, A., Spigel, D., Stassun, K., Street, R., Strolger, LG., Szalay, A., Triafer, J., Troxel, M.A., Turnbull, M., van der Marel, R., van der Linden, A., Wang, Y., Weingberg, D., Williams, B., Windhorster, R., Wolack, E., Wu, H.Y., Yue, J., Zimmerman, N., 2019. The Wide Field Infrared Survey Telescope: 100 Hubbles for the 2020s. arXiv e-prints , arXiv:1902.05669
[3] Araya, M., Osorio, M., Díaz, M., Ponce, C., Villanueva, M., Valenzuela, C., Solar, M., 2018. JOVIAL: Notebook-based astronomical data analysis in the cloud. Astronomy and Computing 25, 110–117. doi:10.1016/j.ascom.2018.09.001
[4] Argudo-Fernández, M., Duarte Puertas, S., Ruiz, J.E., Sabater, J., Verley, S., Bergond, G., 2017. LSSGalPy: Interactive Visualization of the Large-scale Environment Around Galaxies. PASP 129, 058005. doi:10.1088/1538-3873/aa7585
[5] Astropy Collaboration, Price-Whelan, A.M., Sipőcz, B.M., Grieves, H.M., Lim, P.L., Crawford, S.M., Conseil, S., Shupe, D.L., Craig, M.W., Dencheva, N., Ginsburg, A., Vand erPlas, J.T., Bradley, L.D., Pérez-Suárez, D., de Val-Borro, M., Aldcroft, T.L., Cruz, K.L., Robitaille, T.P., Tollerud, E.J., Ardelean, C., Baboj, T., Bach, Y.P., Bachetti, M., Bananov, A.V., Bamber, S.P., Barentsen, G., Barnaby, P., Baumbach, A., Berry, K.L., Biscani, F., Boquien, M., Bostroem, K.A., Bouma, L.G., Branner, G.B., Bray, E.M., Breytenbach, H., Buddeleijmer, H., Burke, D.J., Calderone, G., Can Rodríguez, J.L., Cara, M., Cardoso, J.V.M., Cheedella, S., Copin, Y., Correia, L., Crichton, D., D’Avella, D., Deil, C., Depagne, E., Dietrich, J.P., Donath, A., Droettboom, M., Earl, N., Erben, T., Fabbro, S., Ferreira, L.A., Finethy, T., Fox, R.T., Garrison, L.H., Gibbons, S.L.J., Goldstein, D.A., Gommers, R., Greco, J.P., Greenfield, P., Groener, A.M., Grollier, F., Hagen, A., Hirst, P., Homeier, D., Horton, A.J., Hosseinizadeh, G., Hu, L., Hunkeler, J.S., Ivezic, Ž., Jain, A., Jenness, T., Kanarek, G., Kendrew, S., Kern, N.S., Kerzendorf, W.E., Khvalko, A., King, J., Kirkby, D., Kulkarni, A.M., Kumar, A., Lee, A.L., Lena, D., Littlefair, S.P., Ma, Z., MacLeod, D.M., Mastropietro, M., McCully, C., Montagnac, S., Morris, B.M., Mueller, M., Mumford, S.J., Muna, D., Murphy, N.A., Nelson, S., Nguyen, G.H., Nitnan, J.P., Nöthe, M., Ogaz, S., Oh, S., Parejko, J.K., Parley, N., Pascual, A., Patil, R., Patil, A.A., Plunkett, A.L., Prochaska, J.X., Rastogi, T., Reddy Janga, V., Sabater, J., Sakurikar, P., Seifert, M., Sherbert, L.E., Sherwood-Taylor, H., Shih, A.Y., Sick, J., Silbiger, M.T., Singamanallam, S., Singer, L.P., Sladen, P.H., Sooley, K.A., Sornarajah, S., Streicher, O., Teuben, P., Thomas, S.W., Tremblay, G.R., Turner, J.E.H., Terrón, V., van Kerkwijk, M.H., de la Vega, A., Watkins, L.L., Weaver, B.A., Whitmore, J.B., Wolke, K., Zabalza, V., Astropy Contributors, 2018. The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package. [7] arXiv e-prints, arXiv:1905.02628
[6] Astropy Collaboration, Robitaille, T.P., Terrier, E.J., Greenfield, P., Droettboom, M., Bray, E., Aldcroft, T., Davis, M., Ginsburg, A., Price-Whelan, A.M., Conley, A., Crighton, N., Barbary, K., Muna, D., Ferguson, H., Grollier, F., Parikh, M.M., Nair, P.H., Unther, H.M., Deil, C., Woillez, J., Conseil, S., Kramer, R., Turner, J.E.H., Singer, L., Fox, R., Weaver, B.A., Zabalza, V., Edwards, Z.I., Azalee Bostroem, K., Burke, D.J., Casey, A.R., Crawford, S.M., Dencheva, N.,
[37] Wassong, P., Derriere, S., 2019. The ASTRODEEP Frontier Fields Portal, in: Molinaro, M., Shortridge, K., Pasian, F. (Eds.), Astronomical Data Analysis Software and Systems XXVI, p. 268.

[38] Wells, D.C., Greisen, E.W., Harten, R.H., 1981. FITS - a Flexible Image Transport System. A&AS 44, 363.

[39] Williams, C.C., Curtis-Lake, E., Hainline, K.N., Chevallard, J., Robertson, B.E., Charlot, S., Endsley, R., Stark, D.P., Willmer, C.N.A., Alberts, S., Amorin, R., Arribas, S., Baum, S., Bunker, A., Carniani, S., Crandall, S., Egami, E., Eisenstein, D.J., Ferruit, P., Husemann, B., Maseda, M.V., Maiolino, R., Rawle, T.D., Rieke, M., Smit, R., Tacchella, S., Willott, C.J., 2018. The JWST Extragalactic Mock Catalog: Modeling Galaxy Populations from the UV through the Near-IR over 13 Billion Years of Cosmic History. ApJS 236, 33. doi:10.3847/1538-4365/aabcbb, arXiv:1802.05272

[40] Yu, W., Carrasco Kind, M., Brunner, R.J., 2017. Vizic: A Jupyter-based interactive visualization tool for astronomical catalogs. Astronomy and Computing 20, 128–139. doi:10.1016/j.ascom.2017.06.004, arXiv:1701.01222