Digital Rocks Portal for image curation, characterization, visualization, and transport simulation in porous media

Çınar Turhan[1], Bernard Chang[1], Ali Mohamed[1], Maria Esteva[2], Richard Ketcham[3], James McClure[4] and Maša Prodanović[1,]*

[1]Hildebrand Department of Petroleum and Geosystems Engineering, The University of Texas at Austin, Austin, TX, USA
[2]Texas Advanced Computing Center, The University of Texas at Austin, Austin, TX, USA
[3]Jackson School of Geosciences, The University of Texas at Austin, Austin, TX, USA
[4]National Security Institute, Virginia Tech, VT, USA

Abstract. Modern 3D imaging provides a window to the microstructure of rocks, soil, and other porous materials. Current advances in image characterization, image-based simulation, and machine learning can improve the prediction of transport properties in porous media, and both powerful and easy to use tools and curated datasets are required to facilitate prediction of transport processes. The Digital Rocks Portal (DRP) is an open-science portal for managing, preserving, and analysing porous media imaging data. Implemented within the Texas Advanced Computing Center (TACC) open science research cyberinfrastructure, the portal ascribes to FAIR data principles and connects curated datasets to simulation and analysis tools in a high-performance computing environment (HPC). In this paper we present changes to the portal including new branding, updates in its metadata model, interface design, workflows, and applications. The new set of Python-based applications will be used within the portal in connection to DRP's published datasets or independently as open source published software, and we describe how our new workflows sample "competent" subsets best suited for simulation, machine learning, and visualization tools. In particular, we detail the connection to the open-source lattice Boltzmann method solver LBPM. Finally, we present citation analyses of papers reusing data from DRP and how the findings informed the improvements to the portal's user interface and tools as well as a discussion on the portal's sustainability. The updated portal, which will be renamed Digital Porous Media Portal (DPMP), will help the research community reduce data preparation and post-processing efforts leading to innovations in subsurface porous media and other energy storage materials.

1 Introduction

Computed tomography (CT) and micro-computed tomography (μCT) are now applied routinely to acquire three-dimensional images that reveal the structure of geologic materials on different length scales. Beyond observation and characterization, it allows simulation and thus understanding of mechanisms that directly impact larger scales such as in sedimentary sequences, aquifers, and reservoirs. Experimental imaging has emerged as a leading approach for the study of multiphase flow in porous media [1–3] including entrapped non-wetting phases [4–8], three-fluid-phase systems [9], and the dissolution of an entrapped non-wetting phase [10]. While μCT is one of the most common imaging modalities to produce detailed 3D structure information (and has arguably seen the most recent growth in use), it is far from being the only one [11]. Scanning electron microscopy (SEM) [12] and its recent (pseudo) 3D counterpart focused ion beam (FIB-SEM) [13], confocal microscopy, micromodels [14], magnetic resonance imaging (MRI) [15], particle image velocimetry (PIV) [16], and 4D micro-computed tomography (4D μCT) [17] have all provided valuable insights and detailed textural/elemental information. Techniques such as MRI, PIV, and 4D μCT have the ability to capture dynamic processes.

Each imaging approach is capable of digitally reconstructing objects that range from nanometers to centimeters in size, revealing structure for quantification and investigation of flow and mechanical phenomena. Larger scales of interest with a sufficiently large representative elementary volume (REV) provide monitoring and understanding on practical scales, but often need to be aided by smaller-scale analyses to determine properties such as porosity, permeability, or Young's modulus. As such high-quality imaging is increasingly available, systems to store, retrieve, analyze, and simulate volumetric or large-area images are essential to derive useful information from them. The ability to connect data sources to web-based characterization and simulation resources should be at a researcher's fingertips

* Corresponding author: masha@utexas.edu

as 21st century tools of science, but such capability is currently not available for study of imaged porous media.

Since 2015, the Digital Rocks Portal (DRP) has been a data platform to overcome aforementioned scale and storage problems. Currently it contains 162 porous media datasets and the results from analyses and simulations performed on them. DRP was originally funded by EarthCube [18] and is a member of its Council of Data Facilities. It is also a registered data repository in the Registry of Data Repositories [19], and is a recommended data repository in the Geoscience Data Journal [20]. The objectives of DRP are organizing, preserving, and providing improved access to large scale images and image related data of porous materials, therefore enhancing overall research productivity. The increase in imaging of different porous materials such as energy storage materials like fuel cells and batteries, fibrous materials, and material science media, required a change in the name of the platform. A factor that accelerated this change was the informal feedback gathered from forums and conferences. In the near future, along with ongoing improvements and functional expansion to the portal enabled by an NSF grant RISE-2324786, DRP will be renamed as Digital Porous Media Portal (DPMP) to cover all the mentioned porous materials.

In tandem with the new name, fundamental changes in the portal include a new version of a data model and corresponding metadata fields to represent the new porous media data, and the addition of new analysis and visualization tools that can be conducted within the portal. Analysis tools include geometric characterization of porous materials, heterogeneity analysis, data subset selection for visualization and simulation, and an advanced simulation method, Lattice Boltzmann method (LBM) [21–23]. The latter can be applied to recover a wide range of multi-physics, including the Navier-Stokes equations, multi-phase flow, model diffusion, reactive transport, and heat transfer [24–30]. Since the typical size of 3D images is 500^3 - 2500^3 voxels, parallel implementations and high computational efficiency are essential [30–35] for the future of image-based estimation of porous media properties.

DRP is currently in rebuild phase and will be built using Texas Advanced Computing Center (TACC) infrastructure. Improvements are expected to be completed by August 2024. When finalized, the portal will address the challenge of using high-performance computing (HPC) through a web-based system, as tools will be connected to computational systems of TACC, giving the user an opportunity to access computing power within the portal.

With the changes outlined here, as an open data sharing platform, DRP contributes to open science by:

- Findability of the data by providing unique metadata fields, both machine- and human-readable;
- Accessibility of the data and its metadata through DRP website;
- Organization of the data by linking sample information, digital and analysis datasets, related publications, and related software publications under a main “Dataset” class, and providing an option to connect different datasets using a “Related Datasets” metadata field;
- Permanent accessibility of data by providing a digital object identifier (DOI);
- Facilitate reproducibility and replicability of research through a platform that enables continuous use of datasets;
- Providing open-source tools for characterization, visualization, and simulation through a Digital Porous Media GitHub organization page [36], with current work in progress being containerization of those tools for use within DRP website.

In this paper we describe work to combine data in DRP directly with analysis and HPC simulation, thus creating an open science environment of the 21st century for digital porous media community. We discuss improving the data model inherent to the organization of the portal in Section 2. In Section 3 we provide description of the open science analysis workflows and tools that users can either use as standalone tools or in the next rebuild within the portal applications to directly visualize, analyze or simulate with data, including Lattice Boltzmann methods.

We regularly monitor publications reusing data in the portal through the citations of the datasets using Google Scholar’s alert system, and collect the publications citing data or the portal itself. We periodically review these publications to distinguish use by original vs. other authors, and the type of analyses employed on the data. The result points to trends and informs improvements to the portal user interface, as well as which tools should be connected to the data in the portal. In Section 4 we present the results of one such recent analysis. We summarize our findings in Section 5, and in Section 6 we describe the software stack that will allow all tools and workflows to work directly on the data in the upcoming rebuild.

2 Data Model

The data model guides how data in DRP is organized and described so it can be represented and searched/used in the portal. The schema represents the structure of typical imaging research projects as a system of hierarchical classes with corresponding properties as metadata and the relationships between them. In the user interface the data model is implemented as a form and a data producer creates a “research project” to curate and publish a dataset. Different datasets may have different classes, but all include required ones to assure each is complete and responds to FAIR data standards in terms of metadata, data accessibility, and interoperability. For example, no dataset can be published without sample information.

2.1. Data Model V1.0

The original data model dating from 2015 and referenced as v 1.0 provides the organization and description for the current datasets in the portal (See Fig.1).

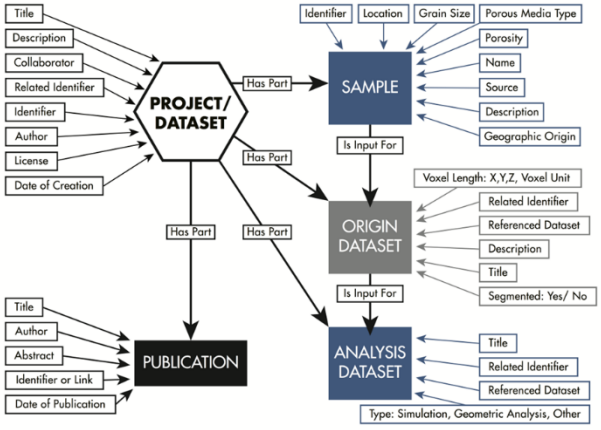

Fig. 1. Data Model (v1.0).

The data model structure has proved to be flexible and appropriate for imaging datasets of geological materials. However, expanding to a wider variety of porous media datasets required a revision of the data model. Therefore, we changed some basic terminology to adjust to old and new media and added new metadata fields.

2.2 Data Model V2.0

The structure of the v2.0 of data model can be seen in Fig. 2, where the green fields indicate newly added metadata fields, and the asterisks indicate the mandatory metadata fields. Below is a list of the new metadata fields.

Under the Dataset class, we added:

i) Related datasets: point to other datasets in the portal that are part of a same research project, derive from an existing dataset or have been reused to produce a new published dataset.
ii) Related software: To note software used to generate, analyze, study the dataset (visualization, simulation, analysis, etc.).

Under Sample we added:

i) A new porous media type, that is called "energy storage media",
ii) New porous media sources are:
    - Natural (earth).
    - Natural (extraterrestrial).
    - Human-made.
    - Computer generated.
iii) Sources origin methods:
    - Collection method for natural sources.
    - Offshore/onshore and water depth for natural (earth) sources.
    - Procedure and equipment for human made sources.
    - Algorithm description for computer generated sources.

Under Digital Dataset we added:

i) Referenced sample.
ii) Dimensionality of the dataset.
iii) Imaging center.
iv) Imaging equipment and model.

Under Analysis Dataset we added:

i) Segmented (yes/no) option.
ii) Referenced sample.
iii) Referenced digital dataset.
iv) New analysis types: machine learning and simulation.

3 Tools and Workflows

The tools discussed here are divided into two sections: visualization and analysis. All the tools presented here are coded using Python programming language except the Lattice Boltzmann simulator, which is discussed in the following sections.

The tools are provided through a GitHub organization named "Digital Porous Media" [36] as stand-alone tools. Once the portal rebuild is finished, we will provide all of these tools also as containerized applications that the portal users can directly use on their data (refer to Section 6 for the Core Experience Portal of TACC).

Any files that are either uploaded or resulted from a containerized application in DRP will show up in the "Data files" area, accessible through the main dashboard of a user account. The files will be stored in the file system on Corral [37], TACC's high performance storage system, in association with the users' portal accounts. Portal accounts on the HPC systems are essentially TACC accounts, which are open to verified academic and industrial users performing research. Usage of TACC's systems is monitored, and any violations are resolved using standard practices that are well established in TACC's usage policies [38]. Data resulting from the analyses can then be published in a new project or downloaded from the user account at their discretion.

3.1 Visualization Tools

We developed visualization tools mainly using a Python library called PyVista [39].

2D visualization tools include the following:

- Histograms tool to view either single or multiple histograms in the same figure;
- Slice plot tool to view various 2D slices of the data;
- GIF tool to generate gifs from 2D slices.

3D visualization tools include the following:

- Orthogonal slices tool that plots orthogonal slices of a dataset (Fig. 3A);
- Isosurface plot (Fig. 3B);
- Glyph plot (Fig. 3C);
- Streamline plot (Fig 3D).

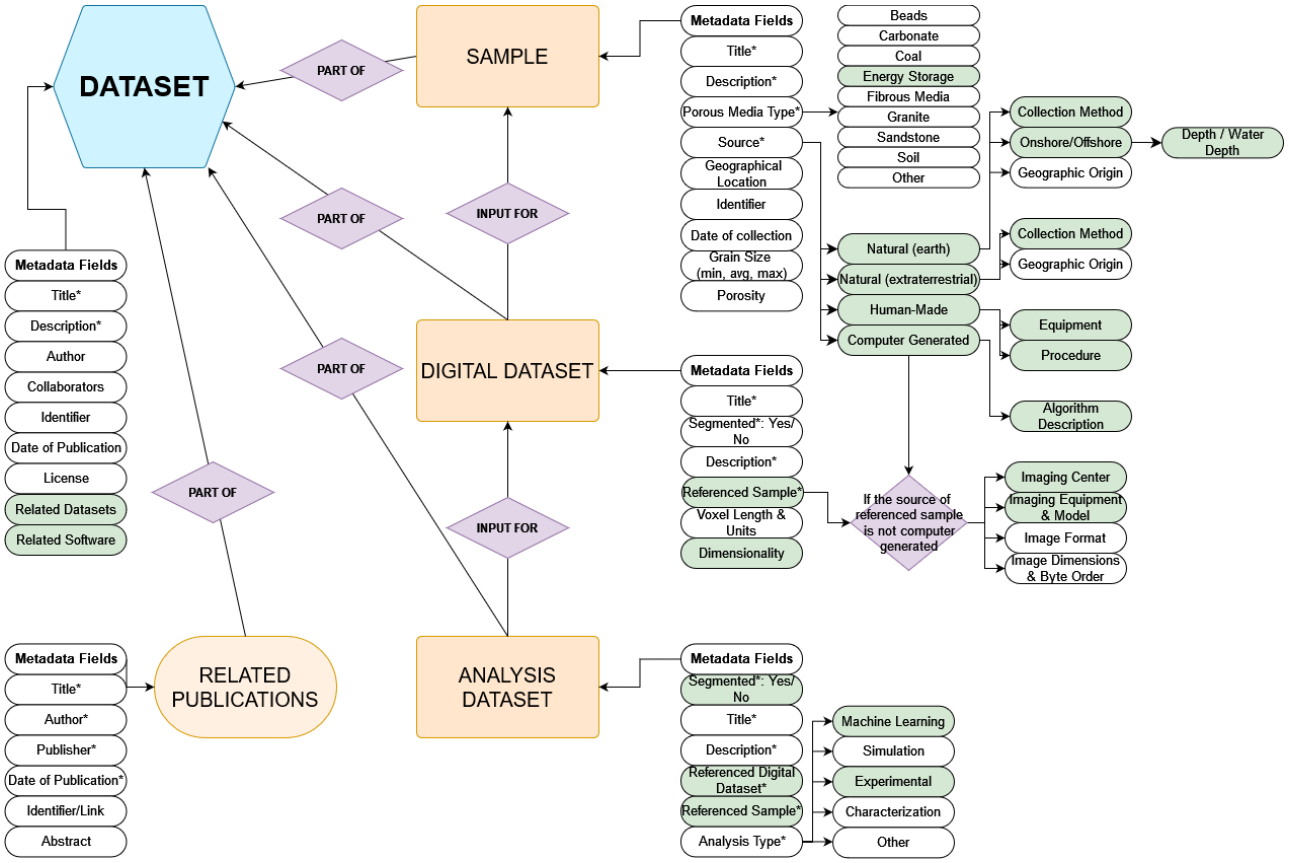

Fig. 2. New Model (v2.0).

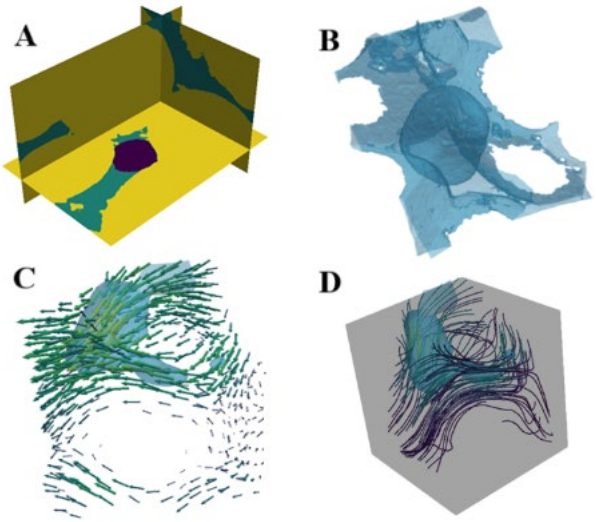

Fig. 3. 3D visualization tools.

3.2 Analysis Tools

Recent developments of DRP includes the development of various analysis tools and Jupyter Notebook workflows. These tools are discussed in detail below.

3.2.1. Minkowski Functionals

The first analysis tool of the portal is a Jupyter Notebook workflow to calculate four of the Minkowski functionals in 3D. These functionals are volume (M0), surface area (M1), integral mean curvature (M2), and Euler characteristic (M3). Each of these functions is implemented using two or more Python packages, mostly using scikit-image [40] and Quantimpy [41]. Their function forms are given with the equations below and exemplified for a single sphere.

$$M_0 = \int_\Omega dV = \frac{4\pi}{3} R^3 \ (\text{Units} \sim l^3) \qquad (1)$$

$$M_1 = \int_{\delta\Omega} dS = 4\pi R^2 \ (\text{Units} \sim l^2) \qquad (2)$$

$$M_2 = \int_{\delta\Omega} \left[\frac{1}{r_1} + \frac{1}{r_2}\right] dS = 4\pi R \ (\text{Units} \sim l^1) \qquad (3)$$

$$M_3 = \int_{\delta\Omega} \frac{1}{r_1 r_2} dS = 4\pi \ (\text{Units} \sim l^0) \qquad (4)$$

3.2.2 Heterogeneity Measurement

Heterogeneity measurement is a Jupyter Notebook workflow that measures whether a rock sample is heterogeneous or homogeneous based on scale-independent porosity variance calculations within a moving window of increasing size in 2D or 3D samples [42]. This classifier was previously used to classify segmented images from the DRP as homogeneous or

heterogeneous with high accuracy, provided no fractures were present. An example of the classification is shown in Fig. 4 and Fig. 5. The starting radius for the moving window for every image is based on the maximum inscribed sphere radius, and the reported radius is relative to the starting radius.

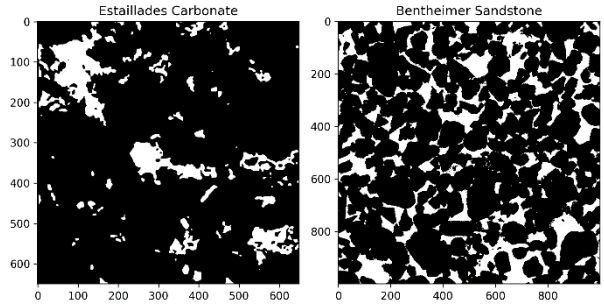

Fig. 4. Cross-sections of two example images, Estaillades Carbonate and Bentheimer Sandstone from Digital Rocks Portal [41, 42]. Solids are shown in black (0), pores in white (1).

3.2.3 Competent Subset Selection

The competent subset selection tool is a tool developed to find a subset in a given 3D sample that is suitable for visualization, simulation, or 3D printing operations. It is implemented for two cases.

The first use case is finding a competent subset for field visualizations. As the scale of velocities is usually in different orders independent of the velocity data source (i.e., experimental or simulation), visualizing velocity fields is tricky. The competent subset selection algorithm for velocity fields bootstraps a subset sized n, and a smaller subset, called n_{sub}. Then, mean, inverse variance, and non-zero velocity components are collected for n, and only the non-zero velocity points are collected in n_{sub}. The n_{sub} part is evaluated to ensure connectivity in the middle part of the visualized subset. In the next step, the harmonic mean of these statistics is calculated, as the aim is to maximize all these values. Based on the collected bootstrapped statistics over a batch, the maximum harmonic mean subset is selected.

The second format of this function is developed for cases without velocity data. Therefore, porosity is the main metric. Under the same workflow, only the parameters are changed. Instead of non-zero velocity points, the number of pixels in the largest connected component is collected, both for n and for n_{sub}. Then, we limited the porosity of the selected competent subset to ± 20% of the original sample porosity to prevent selecting a non-characteristically porous section.

Fig. 6 shows different search spaces for competent subset selection algorithm. Fig. 7 illustrates a competent subset selection for velocity field visualizations.

3.2.4 Interactive Medial Axis Extraction Tool

Medial axis can be used to analyze the geometric structure of the pore space in porous media [45]. The medial axis algorithm starts from a digitized object with the objective to "thin" it down without changing the topology of the object. The algorithms examine voxels (and their neighbors) and removes them if removal does not change topology (e.g., connectivity), until the tinning operation cannot continue further [46]. This medial axis can be used for analyzing connectivity, tortuosity, pore throats, throat barriers, and visualization [47–49].

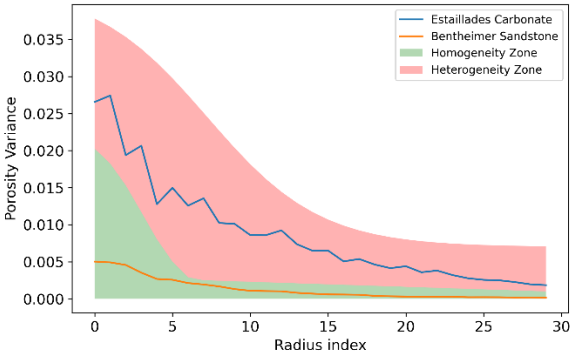

Fig. 5. The heterogeneity assessment curve [42] provides a scale independent measure that quantifies the sample heterogeneity. Here, limestone would be classified as heterogeneous and sandstone as homogeneous. The zones have been determined based on data in DRP in [42].

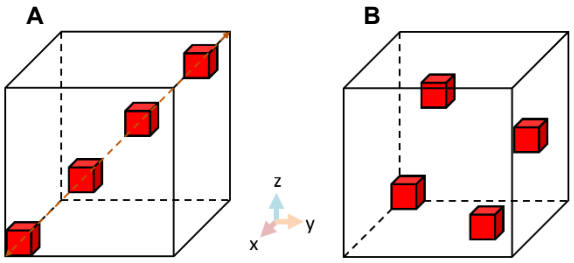

Fig. 6. Diagonal (left) and exhaustive (right) search algorithms for competent sub-set selection.

The developed tool uses scikit-image to extract medial axis, and PyVista to visualize the results. There is an interactive component in this visual, which is a plane. This plane can be moved in one axis of the visual to reveal the medial axis, or to hide it. Fig. 8 illustrates two results of this tool used on a glass bead pack (left) and the Castlegate Sandstone [50] (right), where the red body represents the medial axis.

3.2.5 Image Processing Workflow

An image processing workflow is developed for analyzing hydrate bearing soil images [51]. This workflow is, then, stripped from the dataset and became a standalone tool with the intention of generalizing to other datasets in DRP. It includes beam hardening correction; two image filters, median and anisotropic diffusion; various segmentation methods such as multi-class Otsu thresholding and watershed; and characterization of disconnected components. The workflow is also developed to be suitable for multi-threading to reduce computational time for large datasets.

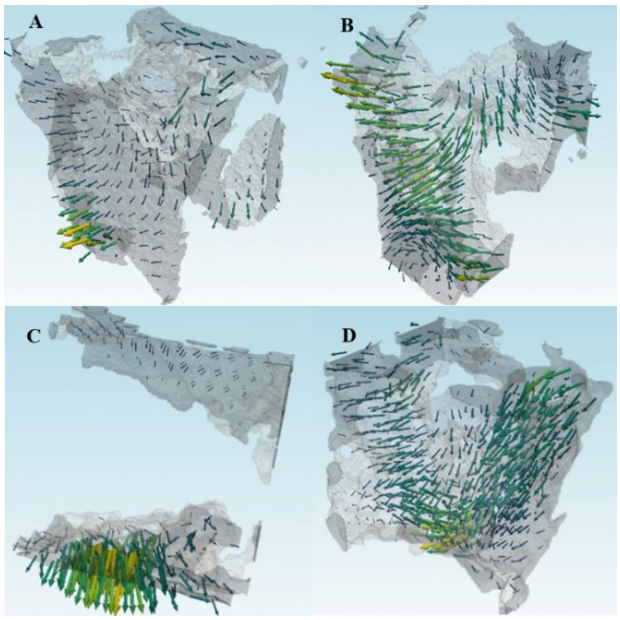

Fig. 7. Images on the left represent random subset selection, and images on the right are the competent subset selections. A&B: subset selections for velocity field visualizations for Bentheimer Sandstone [44]; C&D: same process for Estaillades Carbonate [43].

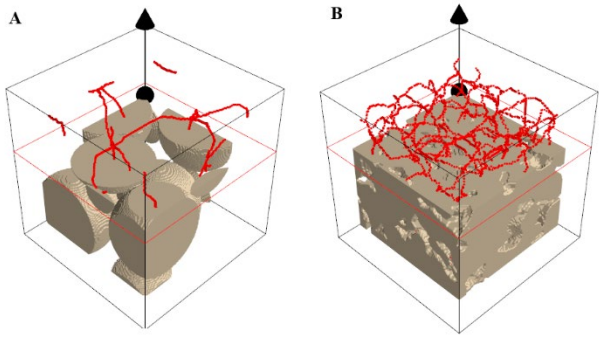

Fig. 8. Medial axis extraction tool results. A: Glass bead pack (subset size is 150^3); B: Castlegate sandstone (subset size is 100^3)

The method to characterize disconnected components was based on the volumes of disconnected components and their aspect ratio, that is, the ratio the of minimum axis length to maximum Feret diameter.

This workflow is planned to be maintained and versioned under the Digital Porous Media GitHub repository [36] and is currently available in there as v1.0 [52] as of April 29, 2024.

3.2.6 Lattice Boltzmann Method Solver (LBPM)

LBPM (Lattice Boltzmann Methods for Porous Media) [53] is an open-source software framework that was specially designed for characterizing various flow processes in digital rock physics workflows. The software package has extensive capabilities in direct simulation of single- and multiphase fluid flow, and electrochemical and ion transport through a variety of porous media. Thus, it can provide many of the image-based properties of porous media desired by the research community, such as relative permeability.

The existing LBPM software framework contains computationally efficient workflows that readily scale from local workstations to exascale supercomputing clusters. It is our motivation to link the DRP datasets to HPC resources in an easy-to-use fashion.

The planned addition of LBPM as a portal application will provide users a streamlined workflow for performing large-scale flow simulations with data hosted on DRP (see more details on future portal applications in Section 6). Users will have direct access to their private workspaces with their unpublished data as well as publicly available dataset for their simulations. They will also be able to launch, monitor, and analyze their simulations using TACC computing resources.

In addition, we developed pyLBPM – a Python-based management interface for the LBPM software that is available under DPM GitHub organization [36]. This wrapper contains modules that will help users install LBPM and its dependencies, design and execute simulation experiments, and track simulation progress. Just as all the previous tools introduced in this paper, LBPM and pyLBPM will be available as an Application through the Portal and work directly with data.

At the time of this writing, the pyLBPM modules are under development. Several features that are already implemented include:

1. Functions to download, configure, and install LBPM and its dependencies directly from a Python command line;
2. CSV, XML, and HDF5 readers specifically designed to ingest simulation outputs and modify simulation inputs using Python;
3. An interactive visualization dashboard to analyze multiphase simulation results (Fig. 9).

The release of the module will also provide a graphical interface to set up simulation experiments, complete with selections and suggestions for required input parameters. It will contain a Python-based launcher to facilitate starting and managing multiple experiments simultaneously. Users that would like to publish the resultant datasets will be able to do so as a new dataset and link it to the one from which one was derived through the Related Dataset metadata element. All datasets are published under a CC BY 3.0 license [54] that asks others to cite the dataset if they use it. DRP publicly distributes and preserves the datasets, but the creators are the ones that obtain credit for their work.

4 Citation Analysis on DRP Data

We conducted a citation analysis to find out how DRP datasets are cited by researchers other than their authors, or in other words data reuse. For the analysis, 22 publications between 2020 and 2023 that cited datasets in

DRP were randomly selected. The study included learning about imaging modality being reused, the domain sciences reusing the data, and whether the study includes image processing or not. This analysis provided good insights into DRP data use and informed the direction for improvements that we are currently implementing, including improving the user interface, developing analysis tools, and adding the new metadata fields.

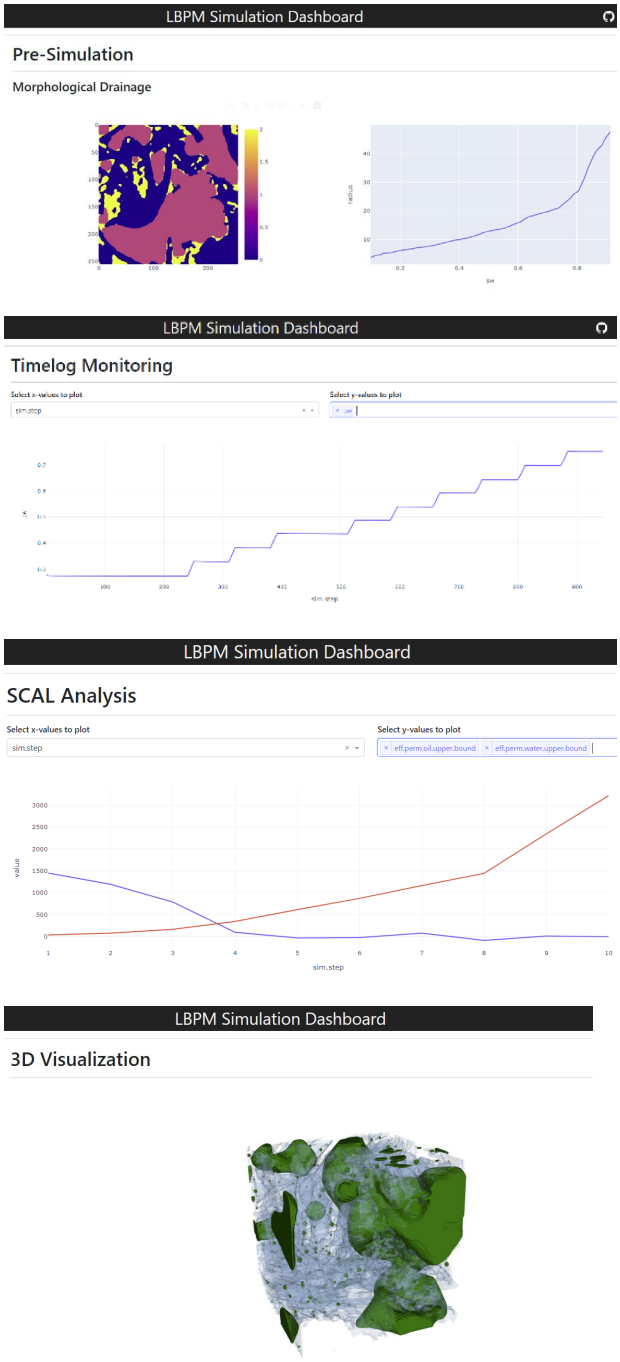

Fig. 9. LBPM Python interface dashboards provide the ability to interact with starting, monitoring and analyzing results of an LBPM simulation. This will become an active application in the Portal with direct link to data. The image reused in this figure is Gambier Limestone from *[50]*.

We observe that the majority (~65%) of the cited data are either micro-CT or CT scans, almost 50% of it being micro-CT. These numbers are followed by X-ray fluorescence and scanning electron microscopy (SEM) datasets. Considering the cost and therefore the availability of X-ray-based industrial devices, this trend is logical. Contrary to SEM images, there is limited sample preparation required for CT scans, which explains this trend. The types of imaging modality results are given in Fig. 10 (left).

55% of all the papers conduct image processing work as shown in Fig 10 (right). The range of image processing methods vary from simple filtering to extensive studies using deep learning. Two papers mention image processing very briefly but don't give method details and thus were not categorized for image processing. Considering the sensitivity of porous media materials to image filtering and segmentation, we expect that researchers conduct their own characterization analyses.

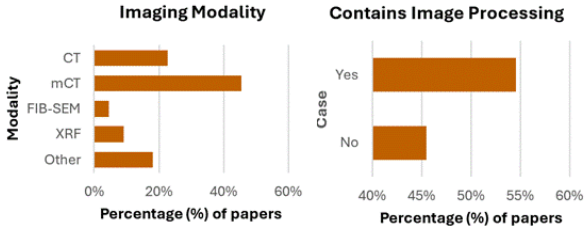

Fig. 10 Imaging modality (left) and image processing methods (right) present in the analyzed papers. Imaging modality is currently not a metadata field, we infer it from the project abstract or the related publication.

As mentioned, a substantial number of the papers focus on porous media characterization. This result is followed by porous media flow with 15%. The remaining papers correspond to different domains, including fracture flow, geomechanics, mineral characterization, porous media reconstruction, and reactive transport (Fig. 11).

We distinguish original use of a dataset, reuse by the authors of the dataset, and reuse by others. If a paper uses a dataset for the first time, and is the uploader to DRP, we call that "original use". If a study uses an existing dataset from DRP, but the authors are the previous publisher of the dataset, this is "reuse by authors". Finally, if the dataset used in a study is an existing dataset, and the authors of the paper are not the publishers of the dataset, this is "reuse by others". The trend for this analysis in Fig. 12 shows a significant percentage of original uploads, with 41%. We also see that reused papers combined have a higher percentage than the original ones. 41% of all the papers are reused by others, and 18% of them are reused by the authors. We can expect to see more original papers as data collection gets easier, and more reuse cases taking advantage of increasing computational algorithms and infrastructure in the following years.

We also collected use cases of these data. This information can be named as analysis type, but it should not be confused with the metadata field "analysis type" of analysis datasets. The data model is only for datasets, whereas this analysis is for the content of the analyzed papers. The use case classes described here are characterization, which represents any kind of digital

analysis to find characteristic properties of porous media; experimental, which corresponds to all laboratory experiments; machine learning (ML), which corresponds to the use of ML and deep learning algorithms; and simulation, which represents any type of simulation study, either numerical or analytical, or both. This "use case" can be named as a "reuse case" if the dataset is reused data from DRP. Of course, some papers might have multiple use cases and the dominant one was counted.

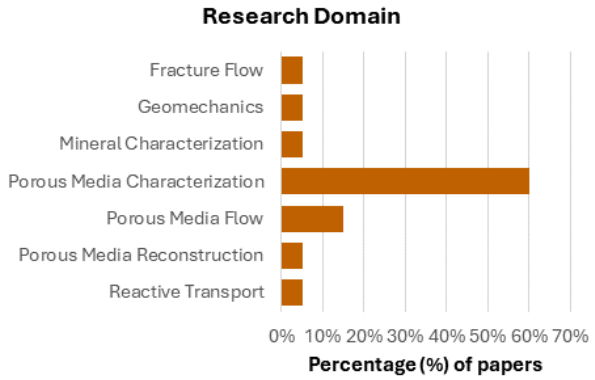

Fig. 11. Research domains of analyzed papers.

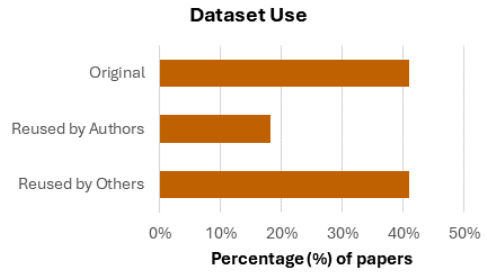

Fig. 12. Dataset use of the analyzed papers.

Figure 13 shows that datasets are mostly used in simulations (35%) followed by ML (26%), experimental (22%), and characterization (17%). The trend can be explained by the increasing availability of computation power around the world. This enables the use of high-performance computing mechanisms even with relatively weaker, local devices. Consequently, in coming years we expect to see more studies employing larger simulations, and larger machine learning/deep learning models.

Figures 11 and 13 indicate the dominant types of analysis are characterization and flow simulation. That informs our choice of tools reviewed in Section 3.2. We further added new analysis types to the data model, experimental and machine learning.

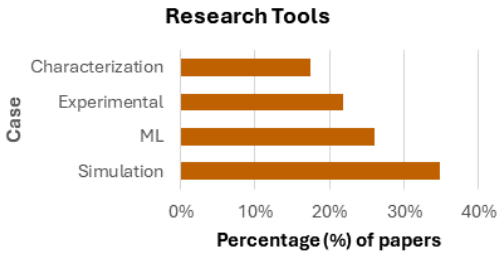

Fig. 13. The research tools utilized in analyzed papers.

The papers and corresponding datasets used in this analysis are in Table 1 in the Appendix.

5 Summary and Conclusion

In this paper, we introduced a new data model for DRP to achieve several improvements over the first model, such as including more porous media types, more metadata fields, and adding clarity to some data classes.

We further introduced new workflows and tools that are planned to be a part of DRP (soon to be renamed DPMP) to directly analyze the datasets through TACC's HPC systems. Moreover, we introduced a Lattice Boltzmann Method Solver (LBPM) that will also be available to users through DRP's new interface.

Finally, we published our analysis results on the publications that use DRP datasets to understand the general trends in imaging modalities, research domains, reuse cases, and dataset uses.

6 Work in Progress and Sustainability

DRP is in the process of rebuilding its infrastructure using TACC's "Core Experience Portal" (CEP) [55]. This change is an important milestone to achieve the sustainability of the portal, as the core infrastructure is maintained by a larger team at TACC, which serves several different portals. This common infrastructure enables DRP to be easily maintained and updated. One of the core capabilities that we are building is connecting data in a portal to the high-performance computing resources that are provided by TACC to UT Austin researchers and beyond. All tools and workflows we report on in this paper are currently standalone tools and will become containerized applications with the newly rebuilt portal. That means that users will be able to submit jobs to queues on various machines (e.g., Lonestar6, which has 560 compute nodes having 256 GB of DRAM, and 85 A100 GPU nodes). This will in turn remove friction from using data in simulation and AI workflows. Fig. 14 shows the portal interface build in progress, where user can go through the "Applications" tab to interact with containerized applications - both remote execution and interactive jobs - and analyze their data on the portal. A zoom in to the tabs in the user interface is given in Fig 15. These applications will be registered with TAPIS [56], an API platform that provides HPC and file management integration to TACC resources, to point at applications metadata on Corral.

We currently charge one-time publication fees for datasets larger than 2GB, and the fee structure was discussed in detail in [57]. The changes to this model are inevitable with the new functionality in simulation and we will have a workshop and community discussion before we establish the new fee structure.

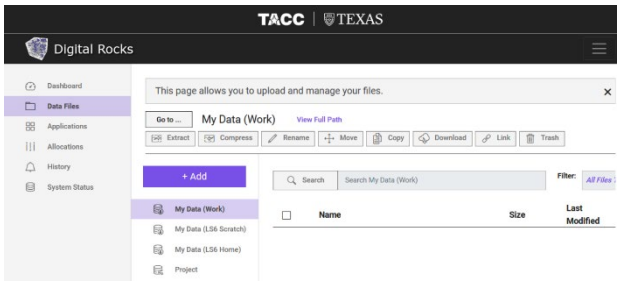

Fig. 14 DRP development stage snapshot as of June 12, 2024.

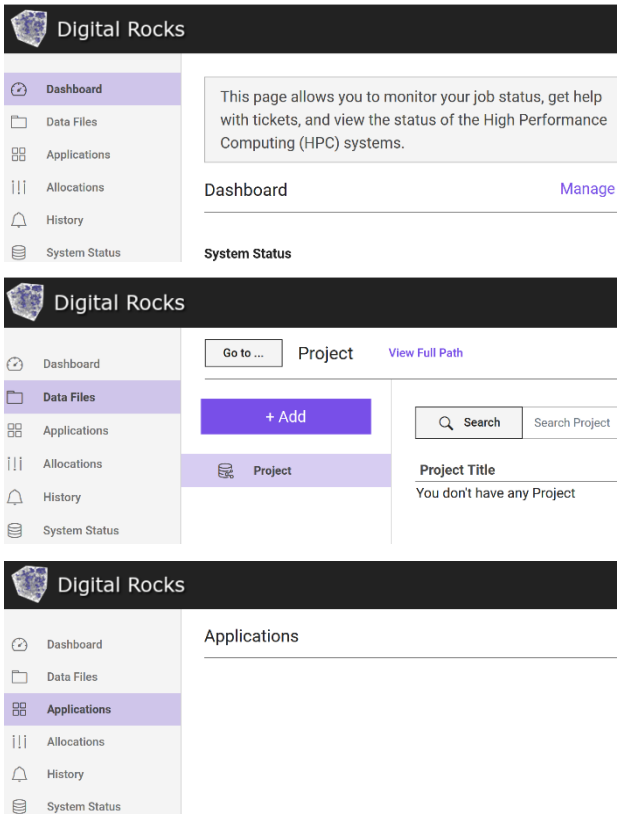

Fig. 15. DRP development stage zoom in snapshots to user dashboard, data files, and applications tabs as of June 12, 2024.

We gratefully acknowledge support from National Science Foundation (NSF GEO OSE grant RISE-2324786) and Digital Porous Media IAP at UT Austin. The portal infrastructure has been implanted on high performance computing resources in Texas Advanced Computing Center.

Table 1. Data use analysis publications list.

Paper	Cited datasets	Imaging Modality	Use*	Reuse Case	Research Domain	I.P.**
[58]	[59]	CT	2	ML	Porous Media Reconstruction	1
[60]	[61]	FIB-SEM; mCT	1	ML	Porous Media Flow	0
[62]	[63]	CT; mCT	2	Characterization	Porous Media Characterization	1
[64]	[65]	CT	2	Simulation	Porous Media Characterization	0
[66]	[59,67]	CT; mCT	2	ML	Porous Media Characterization	0
[68]	[69]	Other	1	Simulation	Porous Media Characterization	0
[70]	[71]	XRF	1	ML	Mineral Characterization	1
[72]	[73–75]	mCT	2	Characterization	Porous Media Characterization	1
[76]	[77]	mCT	1	Experimental	Porous Media Characterization	1
[78]	[67]	mCT	0	ML; Simulation	Porous Media Characterization	1
[79]	[80]	CT	0	Characterization	Porous Media Characterization	1
[81]	[82]	Other	0	Experimental	Porous Media Flow	1
[21]	[22]	mCT	0	Experimental	Fracture Characterization	1
[23]	[71]	XRF	2	Simulation	Fracture Flow	0
[24]	[25]	Other	0	Simulation	Fracture Characterization	0
[26]	[27]	mCT	0	Characterization	Porous Media Characterization	1
[28]	[29]	mCT	0	ML	Porous Media Characterization	1
[30]	[83]	Other	0	Experimental	Porous Media Flow	0
[31]	[32]	mCT	0	Experimental	Porous Media Characterization	1
[33]	[34]	mCT	2	Simulation	Porous Media Characterization	0
[35]	[84]	mCT	2	Simulation	Reactive Transport	0
[85]	[65]	mCT	2	Simulation	Geomechanics	0

* For dataset use, 0,1,2 represents original, reused by the author, and reused by others respectively.

** For image processing (I.P.), 0 and 1 represent the cases where there is not and there is image processing, respectively.

References

1. D. Wildenschild, J. W. Hopmans, M. L. Rivers, and A. J. R. Kent, Vadose Zone J **4**, (2005).
2. D. Wildenschild and A. P. Sheppard, Adv. Water Resour. **51**, 217–246 (2013).
3. T. Bultreys, W. De Boever, and V. Cnudde, Earth-Sci. Rev. **155**, 93–128 (2016).
4. G. Schnaar and M. L. Brusseau, Environ. Sci. Technol. **39**(21), 8403–8410 (2005).
5. M. Prodanović, W. B. Lindquist, and R. S. Seright, J. Colloid Interface Sci. **298**(1), 282–297 (2006).
6. S. Iglauer, A. Paluszny, C. H. Pentland, and M. J. Blunt, Geophys. Res. Lett. **38**(21), L21403 (2011).
7. M. Andrew, B. Bijeljic, and M. J. Blunt, Geophys. Res. Lett. **40**(15), 3915–3918 (2013).
8. M. A. Montoro and F. M. Francisca, Geotech. Test. J. **36**(1), 64–77 (2013).
9. K. Brown, S. Schlüter, A. Sheppard, and D. Wildenschild, J. Microsc. **253**(3), 171–182 (2014).

10. K. C. Carroll and M. L. Brusseau, J. Contam. Hydrol. **106**(1–2), 62–72 (2009).
11. C. J. Werth, C. Zhang, M. L. Brusseau, Oostrom M., and T. Baumann, J. Contam. Hydrol. **113**(1), 1–24 (2010).
12. S. J. B. Reed, Electron Microprobe Analysis and Scanning Electron Microscopy in Geology., 2 edition (Cambridge University Press, 2010).
13. H. Yoon and T. A. Dewers, Geophys. Res. Lett. **40**(16), 4294–4298 (2013).
14. N. K. Karadimitriou and S. M. Hassanizadeh, Vadose Zone J. **11**(3), 0 (2012).
15. S. G. Allen, P. C. L. Stephenson, and J. H. Strange, J. Chem. Phys. **106**(18), 7802 (1997).
16. M. Moroni and J. H. Cushman, Water Resour. Res. **37**(4), 873–884 (2001).
17. S. Youssef, H. Deschamps, J. Dautriat, E. Rosenberg, R. Oughanem, E. Maire, and R. Mokso, *Int. Symp. Soc. Core Anal.* 2013.
18. H. Pampel, P. Vierkant, F. Scholze, R. Bertelmann, M. Kindling, J. Klump, H.-J. Goebelbecker, J. Gundlach, P. Schirmbacher, and U. Dierolf, PLOS ONE **8**(11), e78080 (2013).
19. GFZ German Research Centre For Geosciences, Humboldt-Universität Zu Berlin, Germany Karlsruhe Institute Of Technology (KIT), Purdue University Libraries, R. Bertelmann, M. Buys, H. Cousijn, U. Dierolf, K. Elger, et al, (2013).
20. "Geoscience Data Journal," https://rmets.onlinelibrary.wiley.com/journal/20496060.
21. B. Yang, H. Wang, B. Wang, Z. Shen, Y. Zheng, Z. Jia, and W. Yan, J. Pet. Sci. Eng. **196**, 107682 (2021).
22. B. Yang, "A full-size fracture of tight sandstone induced by water, N2 and CO2" (2020). DRP (DPMP). 10.17612/4W5E-DH86.
23. J. You and K. J. Lee, *SPE Int. Conf. Oilfield Chem.* 2021.
24. K. Sawayama, T. Ishibashi, F. Jiang, T. Tsuji, and Y. Fujimitsu, Rock Mech. Rock Eng. **54**(5), 2145–2164 (2021).
25. K. Sawayama, F. Jiang, and T. Tsuji, "Digitalized natural rock fracture of Inada Granite" (2020). DRP (DPMP). 10.17612/QXSA-TK92.
26. H. Saur, P. Moonen, and C. Aubourg, J. Geophys. Res. Solid Earth **126**(9), e2021JB022025 (2021).
27. H. Saur, C. Aubourg, and P. Moonen, "X-ray micro-CT images of calcareous shale samples" (2021). DRP (DPMP). 10.17612/JEK2-AX94.
28. Y. Niu, Y. D. Wang, P. Mostaghimi, P. Swietojanski, and R. T. Armstrong, Geophys. Res. Lett. **47**(23), e2020GL089029 (2020).
29. Y. Niu, R. Armstrong, and P. Mostaghimi, "Unpaired super-resolution on micro-CT sandstone by using cycle-consistent generative adversarial network" (2020). DRP (DPMP). 10.17612/VZTT-YX38.
30. M. Souzy, H. Lhuissier, Y. Méheust, T. Le Borgne, and B. Metzger, J. Fluid Mech. **891**, A16 (2020).
31. L. E. Dalton, D. Tapriyal, D. Crandall, A. Goodman, F. Shi, and F. Haeri, Transp. Porous Media **133**(1), 71–83 (2020).
32. L. Dalton, "Sessile Drop and Micro-CT Data for Six Sandstone Formations" (2020). DRP (DPMP). 10.17612/GEV9-3M79.
33. M. Wetzel, T. Kempka, and M. Kühn, *Adv. Geosci.* **54**, 33–39 2020.
34. A. Herring, A. Sheppard, M. Turner, and L. Beeching, "Multiphase Flows in Sandstones" (2018). DRP (DPMP). 10.17612/P7MH3M.
35. J.-M. Etancelin, P. Moonen, and P. Poncet, Adv. Water Resour. **146**, 103780 (2020).
36. "Digital Porous Media," https://github.com/digital_porous_media.
37. "Corral," http://tacc.utexas.edu/systems/corral/.
38. "User Policies," http://tacc.utexas.edu/use-tacc/user-policies/.
39. C. B. Sullivan and A. A. Kaszynski, J. Open Source Softw. **4**(37), 1450 (2019).
40. S. Van Der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, and T. Yu, PeerJ **2**, e453 (2014).
41. A. M. P. Boelens and H. A. Tchelepi, SoftwareX **16**, 100823 (2021).
42. A. Mohamed and M. Prodanović, Transp. Porous Media **150**(2), 257–284 (2023).
43. B. P. Muljadi, "Estaillades Carbonate" (2015). DRP (DPMP). 10.17612/P73W2C.
44. B. P. Muljadi, "Bentheimer Sandstone" (2015). DRP (DPMP). 10.17612/P77P49.
45. W. B. Lindquist, S.-M. Lee, D. A. Coker, K. W. Jones, and P. Spanne, J. Geophys. Res. Solid Earth **101**(B4), 8297–8310 (1996).
46. T. C. Lee, R. L. Kashyap, and C. N. Chu, CVGIP Graph. Models Image Process. **56**(6), 462–478 (1994).
47. A. Mehmani, R. Verma, and M. Prodanović, Mar. Pet. Geol. **114**, 104141 (2020).
48. M. Prodanović, W. B. Lindquist, and R. S. Seright, J. Colloid Interface Sci. **298**(1), 282–297 (2006).
49. M. Prodanović, S. L. Bryant, and J. S. Davis, Transp. Porous Media **96**(1), 39–62 (2013).
50. A. Sheppard and G. Schroeder-Turk, "Network Generation Comparison Forum" (2015). DRP (DPMP). 10.17612/P7059V.

51. C. Turhan, "Towards Scalable Data Model for Curation and Reusable Workflows for Porous Media Image Analysis," The University of Texas at Austin (2024).
52. C. Turhan and M. Prodanovic, (2024).
53. J. E. McClure, Z. Li, M. Berrill, and T. Ramstad, Comput. Geosci. **25**(3), 871–895 (2021).
54. "CC BY 3.0 US Deed | Attribution 3.0 United States | Creative Commons," https://creativecommons.org/licenses/by/3.0/us/deed.en#:~:text=Attribution%20%E2%80%94%20You%20must%20give%20appropriate,endorses%20you%20or%20your%20use.
55. "Core Experience Portal," https://cep.tacc.utexas.edu/.
56. J. Stubbs, R. Cardone, M. Packard, A. Jamthe, S. Padhy, S. Terry, J. Looney, J. Meiring, S. Black, M. Dahan, S. Cleveland, and G. Jacobs, *Adv. Inf. Commun.* 878–900 2021.
57. M. Prodanović, M. Esteva, J. McClure, B. C. Chang, J. E. Santos, A. Radhakrishnan, A. Singh, and H. Khan, E3S Web Conf. **367**, 01010 (2023).
58. Y. Li, G. Han, and P. Jian, Geoenergy Sci. Eng. **228**, 212005 (2023).
59. R. Neumann, M. Andreeta, and E. Lucas-Oliveira, "11 Sandstones: raw, filtered and segmented data" (2020). DRP (DPMP). 10.17612/F4H1-W124.
60. J. E. Santos, A. Marcato, Q. Kang, M. Mehana, D. O'Malley, H. Viswanathan, and N. Lubbers, (2023).
61. J. E. Santos, B. Chang, Gigliotti, Alex, Kang, Qinjun, Lubbers, Nicholas, Viswanathan, Hari, and Prodanovic, Masa, "3D Dataset of Simulations" (2021). DRP (DPMP). 10.17612/93PD-Y471.
62. L. da S. Bomfim, G. D. Avansi, A. C. Vidal, and H. Pedrini, Rev. Bras. Cartogr. **75**, (2023).
63. D. N. Espinoza, "Naturally fractured coal sample" (2015). DRP (DPMP). 10.17612/P7QP4N.
64. F. Az-Zahra and I. A. Dharmawan, CFD Lett. **15**(12), 1–18 (2023).
65. Z. Karpyn, C. Landry, and M. Prodanovic, "Induced rough fracture in Berea sandstone core" (2016). DRP (DPMP). 10.17612/P7J012.
66. X. Zou, C. He, W. Guan, Y. Zhou, H. Zhao, and M. Cai, Energy **285**, 129512 (2023).
67. N. Alqahtani, P. Mostaghimi, and R. Armstrong, "A Multi-Resolution Complex Carbonates Micro-CT Dataset (MRCCM)" (2021). DRP (DPMP). 10.17612/3T36-Q704.
68. K. Sawayama, T. Ishibashi, F. Jiang, and T. Tsuji, Geophys. Res. Lett. **50**(20), e2023GL104418 (2023).
69. K. Sawayama, "15 fractures of granite" (2023). DRP (DPMP). 10.17612/7ffv-c780.
70. J. J. Kim, F. T. Ling, D. A. Plattenberger, A. F. Clarens, and C. A. Peters, Appl. Geochem. **136**, 105162 (2022).
71. C. Peters and J. Kim, "Eagle Ford Shale: Synchrotron-Based Element and Mineral Maps" (2020). DRP (DPMP). 10.17612/T3A6-6356.
72. N. Castillo, J. R. Campos, R. Palomino, M. A. Caja, and C. Santos, **2022**(1), 1–5 2022.
73. M. Andrew, B. Bijeljic, and Blunt, Martin, "Doddington Sandstone" (2020). DRP (DPMP). 10.17612/NHDA-WZ02.
74. A. H. Kohanpur, V. Albert, and D. Crandall, "Micro-CT images of a heterogeneous Mt. Simon sandstone sample" (2019). DRP (DPMP). 10.17612/1DVH-1N64.
75. T. Ramstad and A. Kristoffersen, "Bentheimer Micro_CT with Waterflood" (2018). DRP (DPMP). 10.17612/P7795W.
76. S. Wang, L. C. Ruspini, P.-E. Øren, S. Van Offenwert, and T. Bultreys, Water Resour. Res. **58**(1), e2021WR030870 (2022).
77. S. Wang, T. Bultreys, S. Van Offenwert, and L. Ruspini, "Dataset for unsteady-state capillary drainage experiment on Estaillades carbonate" (2022). DRP (DPMP). 10.17612/6RTT-5W16.
78. N. J. Alqahtani, Y. Niu, Y. D. Wang, T. Chung, Z. Lanetc, A. Zhuravljov, R. T. Armstrong, and P. Mostaghimi, Transp. Porous Media **143**(2), 497–525 (2022).
79. E. J. Goldfarb, K. Ikeda, R. A. Ketcham, M. Prodanović, and N. Tisato, Comput. Geosci. **159**, 105008 (2022).
80. E. Goldfarb, M. Prodanovic, R. Ketcham, N. Tisato, and K. Ikeda, "Targeted CT: Predictive Digital Rock Physics Models Without Segmentation" (2020). DRP (DPMP). 10.17612/9X96-AC88.
81. N. Wang, Y. Liu, L. Cha, M. T. Balhoff, and M. Prodanovic, SPE J. **27**(01), 753–770 (2022).
82. N. Wang, Y. Liu, L. Cha, M. Prodanovic, and M. Balhoff, "Ferrofluid displacement in a simple 2.5D micromodel" (2021). DRP (DPMP). 10.17612/2XZJ-FD68.
83. M. Souzy, H. Lhuissier, Y. Méheust, T. Le Borgne, and B. Metzger, "Experimental 3D Velocity Field in Random Sphere Packing" (2020). DRP (DPMP). 10.17612/HDP8-0149.
84. L. Ferreira, R. Surmas, M. Silva, and R. Peçanha, "Carbonates: Porosity and permeability voxel to voxel" (2020). DRP (DPMP). 10.17612/V09Y-AW80.
85. A. Rodríguez de Castro, A. Ahmadi-Sénichault, and A. Omari, Adv. Water Resour. **146**, 103794 (2020).