Submitted on: 27.06.2022

Online Data Research Repositories and Digital Scholarly Ecosystems: From Research Data and Datasets to Artificial Intelligence and Discovery

Raymond Uzwyshyn
Division of Library Collections and Digital Services
Texas State University, San Marcos, USA.
E-mail address: ruzwyshyn@gmail.com

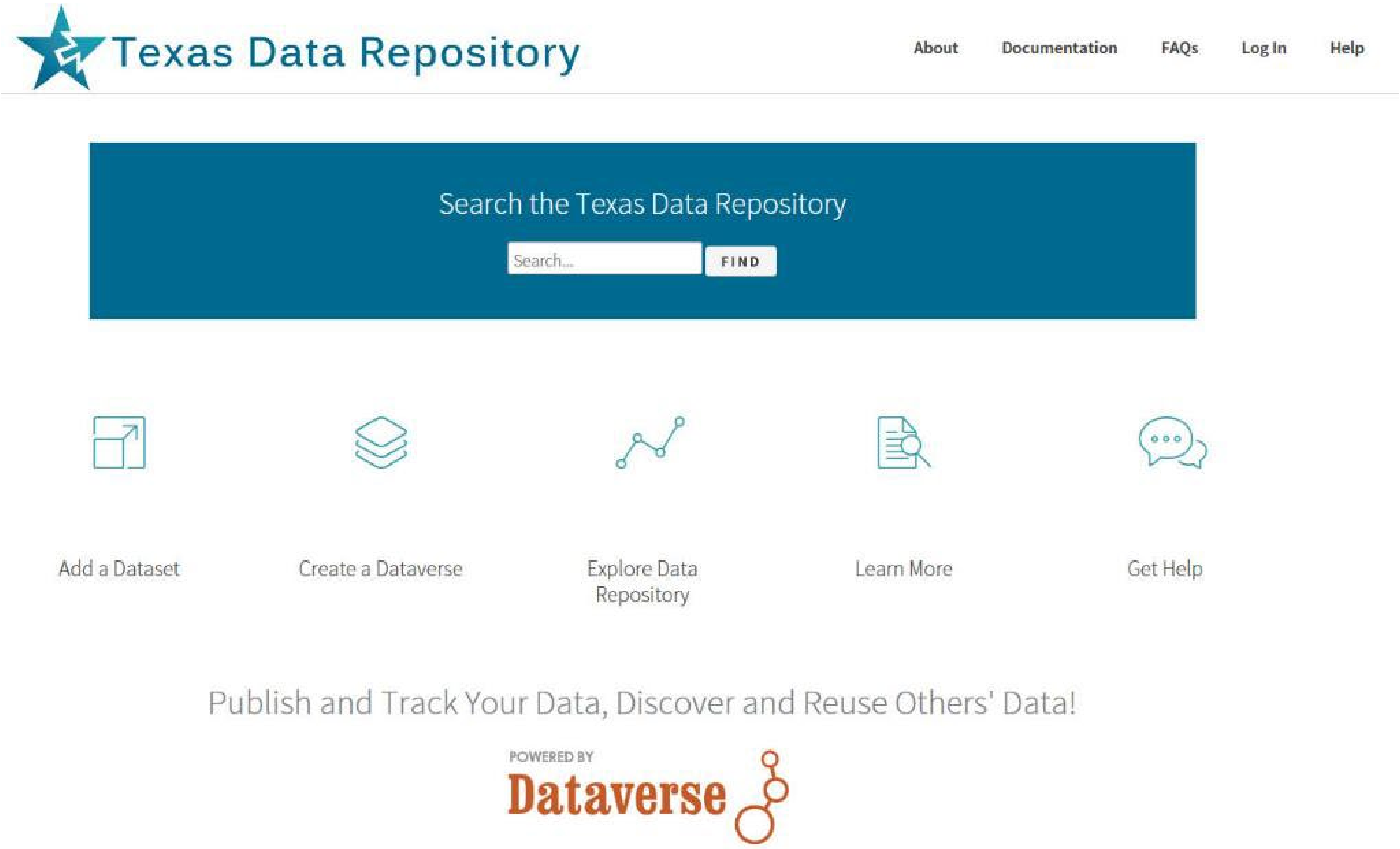
ABSTRACT:

Online Data research repositories are currently being leveraged to accelerate global research, promote international collaboration, and innovate on levels previously thought impossible. Research data repositories may also link data to further content from online publications and other digital communication and aggregation tools.

This article pragmatically overviews such a data and content-centered ecosystem at Texas State University Libraries in the United States. The research then goes on to speak about the ecosystem's next level of planning and construction involving both bigger data possibilities and AI infrastructures. The research uses examples of recent digitized medical image datasets and Deep Learning/Neural Net areas of Artificial Intelligence for global open science possibilities. These methodologies show large promise in making very good use of online open data repositories, digital library ecosystems and online datasets.

An online data-centered research ecosystem accelerates open science, research and discovery on global levels. This open-source ecosystem and software infrastructure may be easily replicated by research institutions. Creating open online data infrastructures for research communities enables future global data and research aggregation, collaboration and the advancement of science, the academic research cycle and new discovery for present and future networked global levels.

Keywords: Artificial Intelligence, Neural Nets, Libraries, Big Data, Data Research Repositories, Online Research Ecosystems



Texas Data Repository, http://data.tdl.org

1 INTRODUCTION

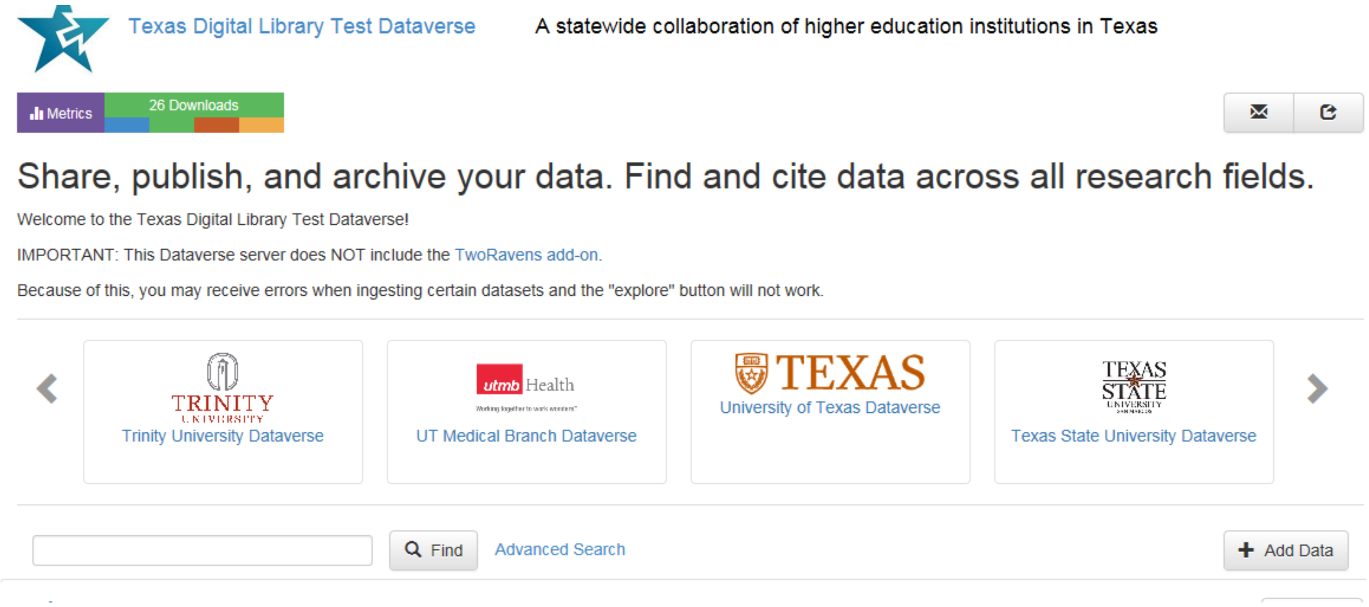
The path from data and experimental research data to online data repositories, research ecosystems and artificial intelligence discovery is not overly clear. This research overviews necessary infrastructures from an online research data repository and digital scholarly ecosystem within currently globally occurring Artificial Intelligence research discovery and insight to present better clarity towards these present necessities.

This paper uses the example of Texas State University Libraries research data infrastructure to overview rudiments of an online data repository and the utility for placing a data repository within a larger scholarly research ecosystem. Current possibilities of online data allow large discovery within Artificial Intelligence, particularly Deep Learning and Neural Nets. This research illustrates this with two online medical related image dataset examples to foreground the importance of online data repository ecosystems towards open science and, particularly, Artificial Intelligence and new discovery. The first example is from a larger scientific deep learning neural net discovery towards cancer detection utilizing object recognition and big data for machine learning and neural net training from a US university (Stanford). The second example builds on the first models and methodology utilizing an undergraduate student theses from BRAC University from Dhaka Bangladesh. Examples begin with Stanford's large AI neural net model and then focus on data repository and ecosystem methdologies which continue to focus Stanford's

AI neural net research with other smaller, online, openly available datasets. Both examples give compelling evidence illustrating the larger value of online open data research respositories and data-centered scholarly ecosystems towards the future progress of science and discovery in our new millenia. New potential for open science is enabled by the recent constellations of global networks, algorithmic Artificial Intelligence Neural Network Deep Learning models, online data research repositiories, ecosystems and increasing computing processing power and storage. Basic rudiments of an online data research repository, scholarly research ecosystem are outlined. Examples are then utilized to show to how these new infrastructures may be used to enable the new potential for AI for scientific discovery in the 21st century.

2 WHAT IS AN ONLINE RESEARCH REPOSITORY?

An online data research repository allows one to share, publish and archive a researcher's data. It is at once a platform to manage a researcher's and institution's data and metadata, a permalinking strategy for Data Citation, a way to manage increasingly mandated large grant compliance and an efficacious data archiving and sharing strategy.



Texas Digital Library Test Dataverse

A statewide collaboration of higher education institutions in Texas

Share, publish, and archive your data. Find and cite data across all research fields.

Welcome to the Texas Digital Library Test Dataverse!

IMPORTANT: This Dataverse server does NOT include the TwoRavens add-on.

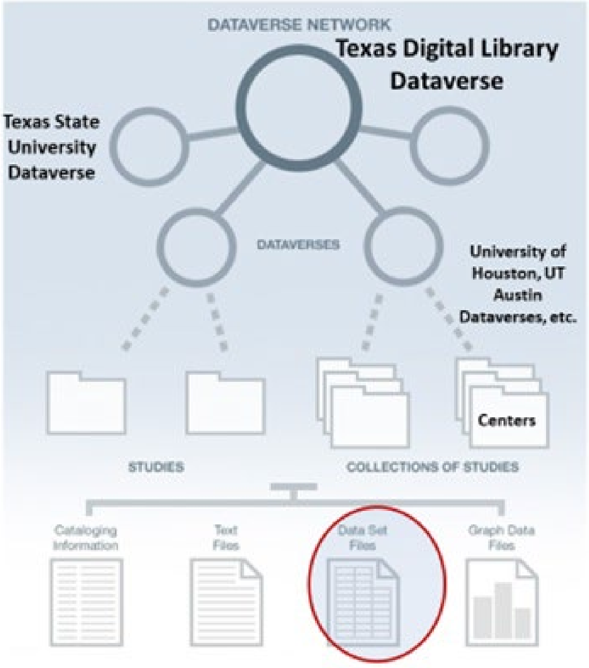
Because of this, you may receive errors when ingesting certain datasets and the "explore" button will not work.

Metrics
26 Downloads
Trinity University Dataverse
UT Medical Branch Dataverse
University of Texas Dataverse
Texas State University Dataverse
Find
Advanced Search
Add Data

Texas Data Research Repository, https://dataverse.tdl.org

The Texas Data Repository is a good example of a consortial data repository and utilizes Harvard's open source Dataverse Software customized towards a consortial multi-university strategy.[1] The Texas Data Repository aggregates various individual university's data for the search and retrieval and can be configured as a single instance for searching or to search across the entire group of institutions.

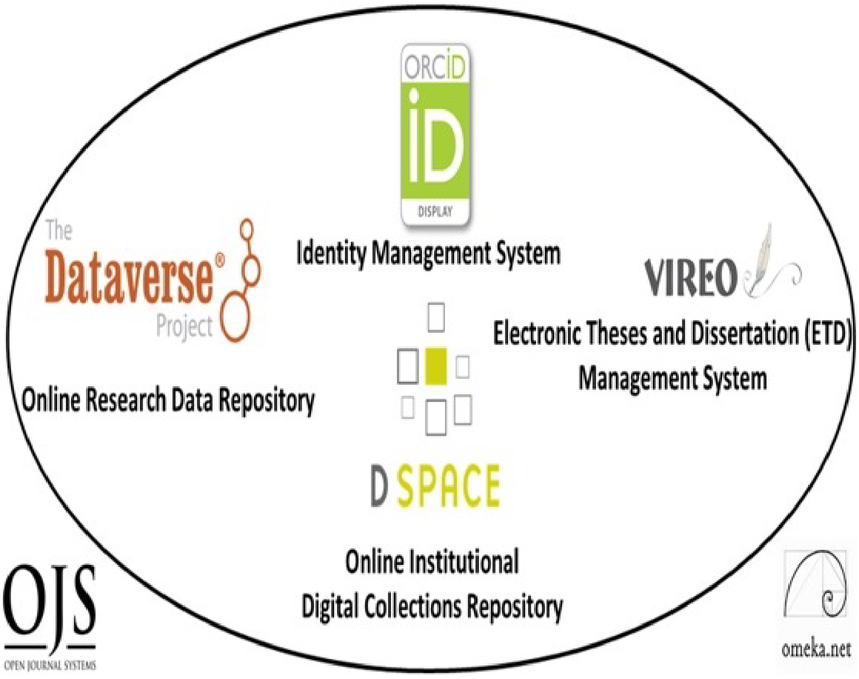
[1] Also See Uzwyshyn, Online Data Repositories (2016). https://www.researchgate.net/publication/304780954_Online_Research_Data_Repositories_the_What_When_Why_and_How



Texas Data Repository Consortial Architecture

3 DIGITAL SCHOLARSHIP ECOSYSTEMS

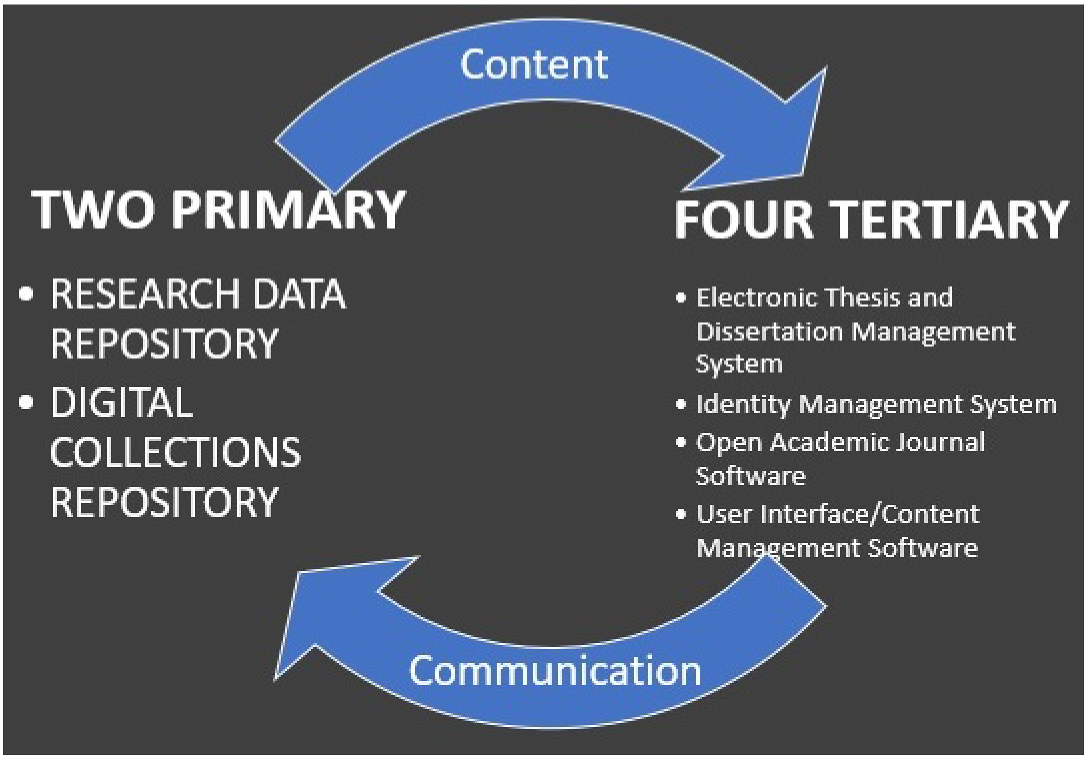
A Digital Repository may also be placed within a larger digital scholarship ecosystem which enables a wider horizon of content and global communications network communication.


ORCiD
iD
DISPLAY
Identity Management System
The Dataverse® Project
Online Research Data Repository
VIREO
Electronic Theses and Dissertation (ETD) Management System
D SPACE
Online Institutional Digital Collections Repository
OJS
OPEN JOURNAL SYSTEMS
omeka.net

The Texas State Digital Scholarship Research Ecosystem Consists of Six Components

The Texas State Digital Scholarship ecosystem utilizes the well-known open-source repository software, Dspace, for the university's digital collections repository. Four other tertiary components are also utilized by researchers to better enable online global communication and network possibilities. These four applications are an online electronic theses and dissertation management system, ETD System (VIREO), identity management system (ORCID), open academic journal system software (OJS3) and user interface content management software (OMEKA). Together, these function as a unified digital scholarship ecosystem.[2]

This ecosystem allows for great facility in enabling data-centered methodologies and continuing to build on strong foundations and providing foundational training data for later AI pathways that may be needed online.


Content
TWO PRIMARY
RESEARCH DATA REPOSITORY
DIGITAL COLLECTIONS REPOSITORY
FOUR TERTIARY
Electronic Thesis and Dissertation Management System
Identity Management System
Open Academic Journal Software
User Interface/Content Management Software
Communication

A Digital Scholarship Research Ecosystem is enabled by both Content and Communication Components

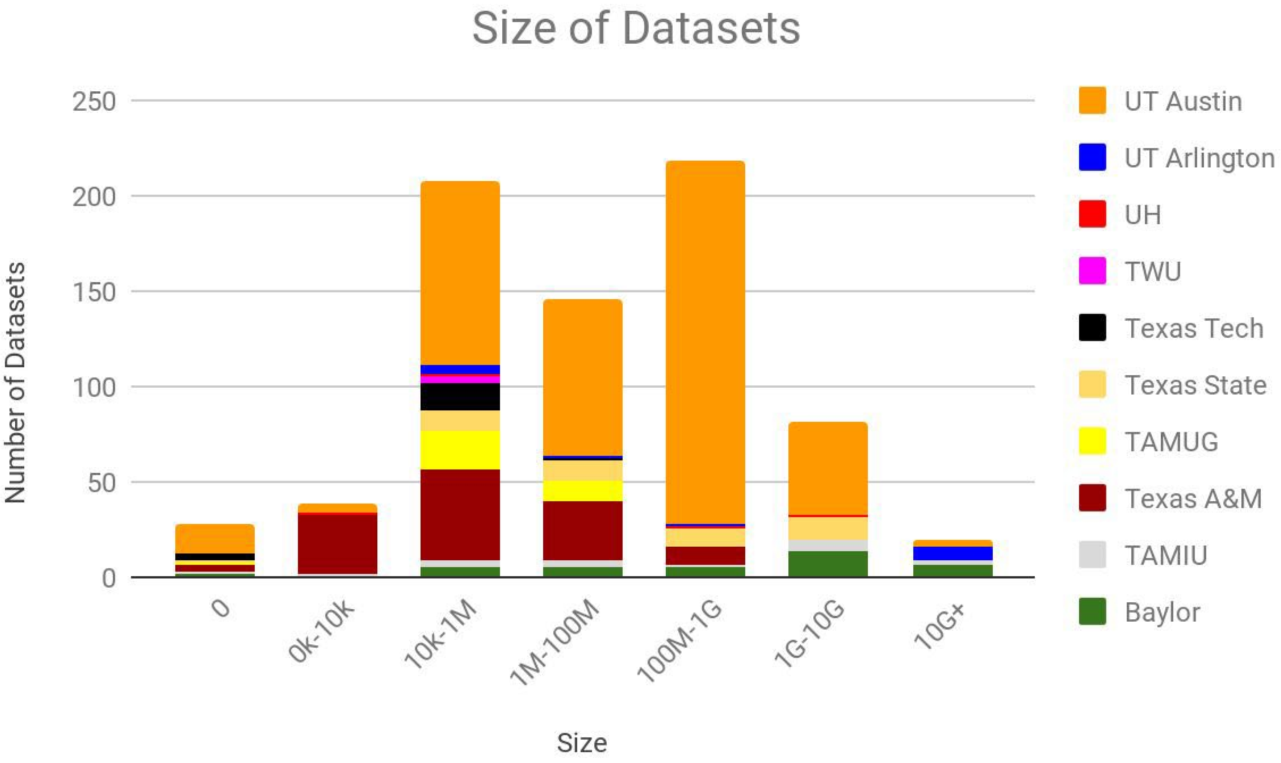
The general common characteristics for such a digital system are open-source software, active developer communities and customizable communication and content repository components. The open-source software allows later customizability and connection between components. Active developer communities for the software enable a lively exchange of new possibilities with regards to innovation. The open-source code allows bridges among systems so that the sum of the system's capabilities exceeds separate parts. Collocating open-source digital components in a networked research ecosystem also enables larger connections network effects of which many possibilities are still untapped.

Together, these digital ecosystem components also enable a larger academic research cycle from original search and retrieval of data and content to gathering and analysis of data, to later writing, publishing and sharing online.

[2] See Uzwyshyn, 2020. Available at: https://www.researchgate.net/publication/336923249_Developing_an_Open_Source_Digital_Scholarship_Ecosystem

4 DATA, DATASETS AND BIG DATA

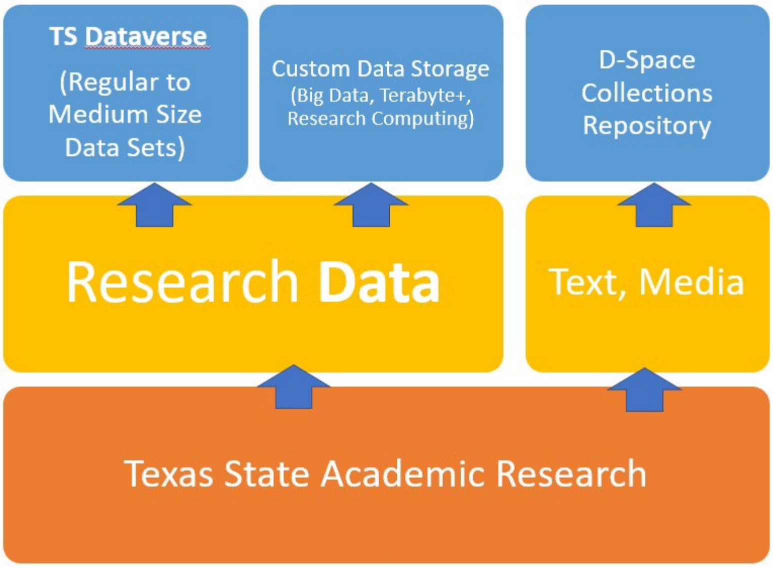
Data comes in a variety of file types, formats, media, and sizes. For AI and particularly recent Deep Learning both larger labelled and unlabelled datasets become important for machine training. Within information science, metadata becomes key and information science's disciplinary schema systems are very useful. One size also does not fit all for various data research repository project needs and there are many types of sizes for data projects and repositories. The Texas Data Repository utilizing Dataverse can upload currently up to 4GB data for individual files and 10GB Datasets. This may not seem large currently in terms of some recent examples of mammoth natural language processing or image/video model huge datasets being trained by Google's DeepMind or Microsoft's Open AI (see Mitchell, 2022) which utilize Terabytes and Exabytes of data. These serve the needs of most academic researchers and have served researchers well for the last five years (2017-2022).



Sizes of Texas Data Research Repository Datasets (See Waugh, 2020)

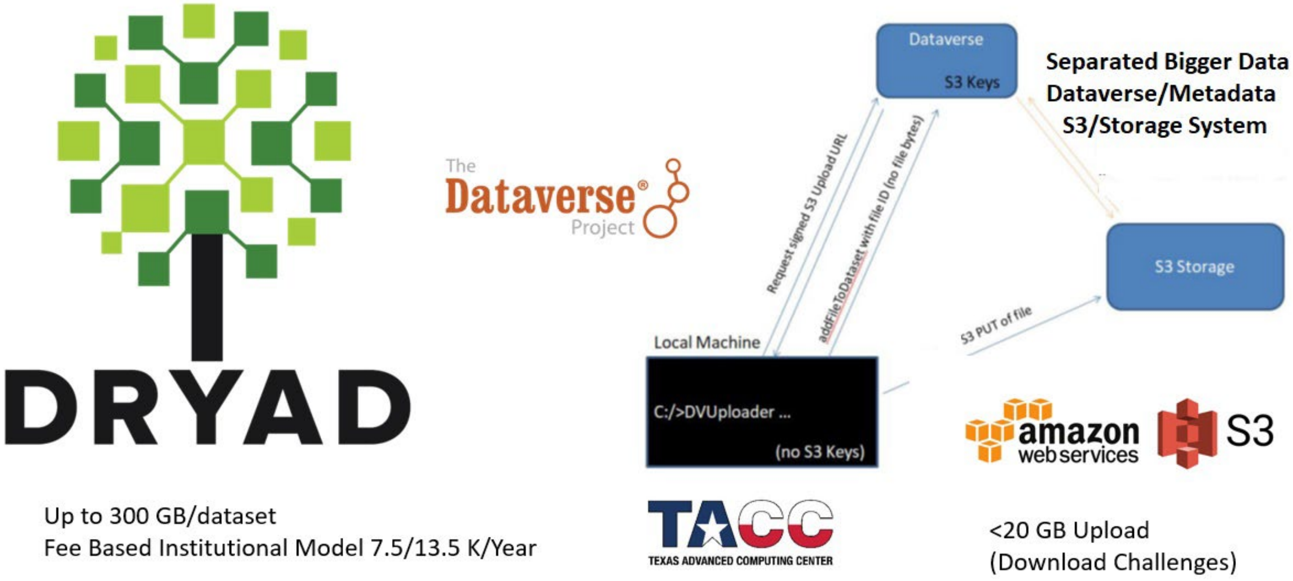
Most researchers collected datasets for upload are still in the 1< <1000 MB range. Currently, there is the growing recognition by researchers that 'bigger' data repositories are needed now beginning in the Giga and Terabyte ranges and readying for the next phase of build. Many of these researchers are also more on the leading edge with specialized media or GIS datasets. In these cases, for larger, bigger and custom data storage it is still not yet overly feasible to place these huge datasets online, especially those in the Terabyte or Exabyte range. Preferably, these are placed with university research computing data centers or the local area supercomputing center for custom data storage should these needs arise.

This type of storage is usually worked out by researchers in preliminary researcher grant applications expecting this level of data storage needed for research work and grant applications.


TS Dataverse
(Regular to Medium Size Data Sets)
Custom Data Storage
(Big Data, Terabyte+, Research Computing)
D-Space Collections Repository
Research Data
Text, Media
Texas State Academic Research

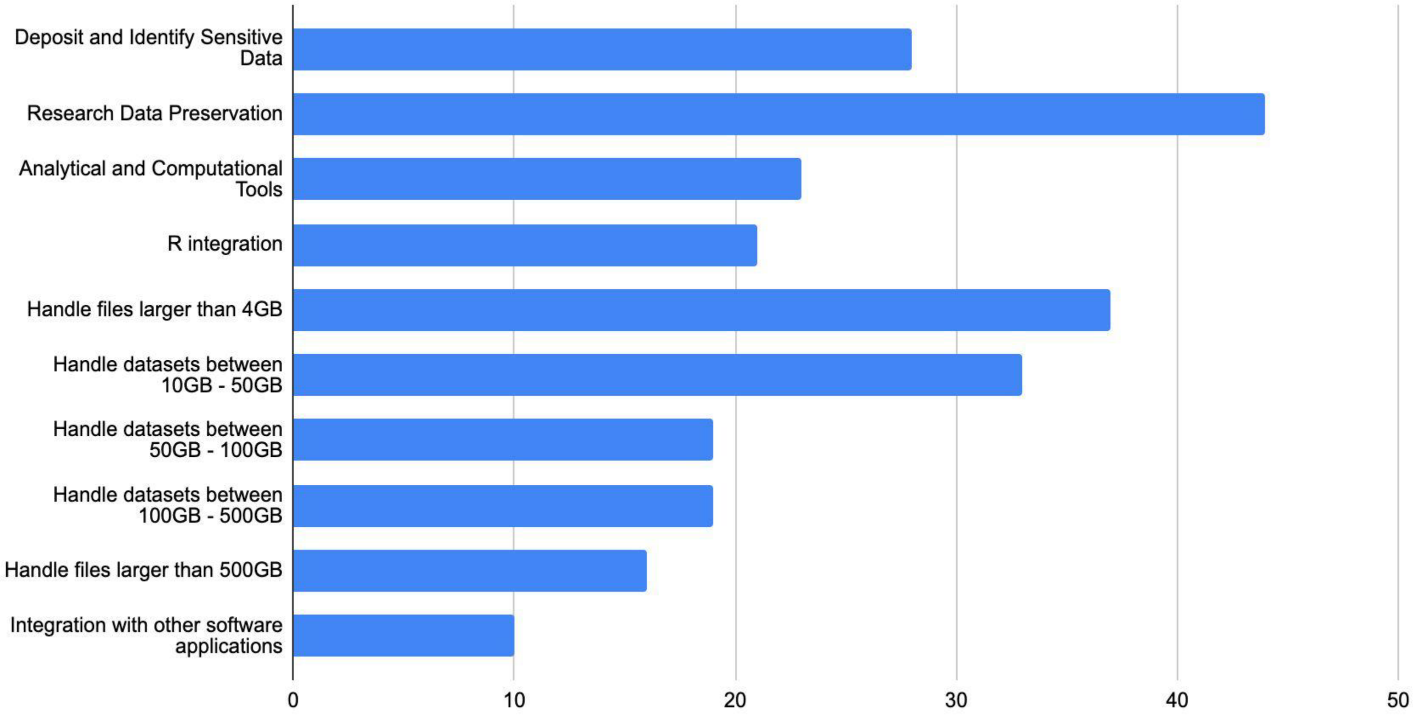
Texas State Universities Big Data Storage Model

Currently, beyond the few custom big data storage needs the requests for very ‘big data’ (Terabytes, Exabyte storage) are still few but these requests are increasing. In this regard, Texas State University Libraries have been exploring various ‘bigger data options and beta prototypes (2020-2022) ranging from 20GB expansions (Amazon Web Services S3 storage) and the Texas Advanced Computing Center) with separated metadata/storage pointer systems to more fee-based institutional model up to 300GB/dataset (Data Dryad).



Beta Prototyping Bigger Data Online Texas Data Repository Architectures, 2020-2022, TACC: https://www.tacc.utexas.edu/, Data Dryad https://datadryad.org/stash

Currently, 'Big Data' (Exabyte, Terabyte) is not at the top of the list of new data research repositories feature set requests that most researchers would like to see. Higher on this list of new features is long term research data digital preservation[3]. Also ranking high, is handling slightly larger data files (4-10 GB range) and datasets between the 10-50GB range as well as being able to safely deposit and clean sensitive data (i.e., Medical related etc., see data survey below). Greater support for analytical and computational tools also comes high on the list. These tool and data literacy requests, ranging from data analytics and visualization, help to enable researchers from non-Computer Science disciplines towards new AI methodologies such as those being forwarded currently through neural net and deep learning methodologies.



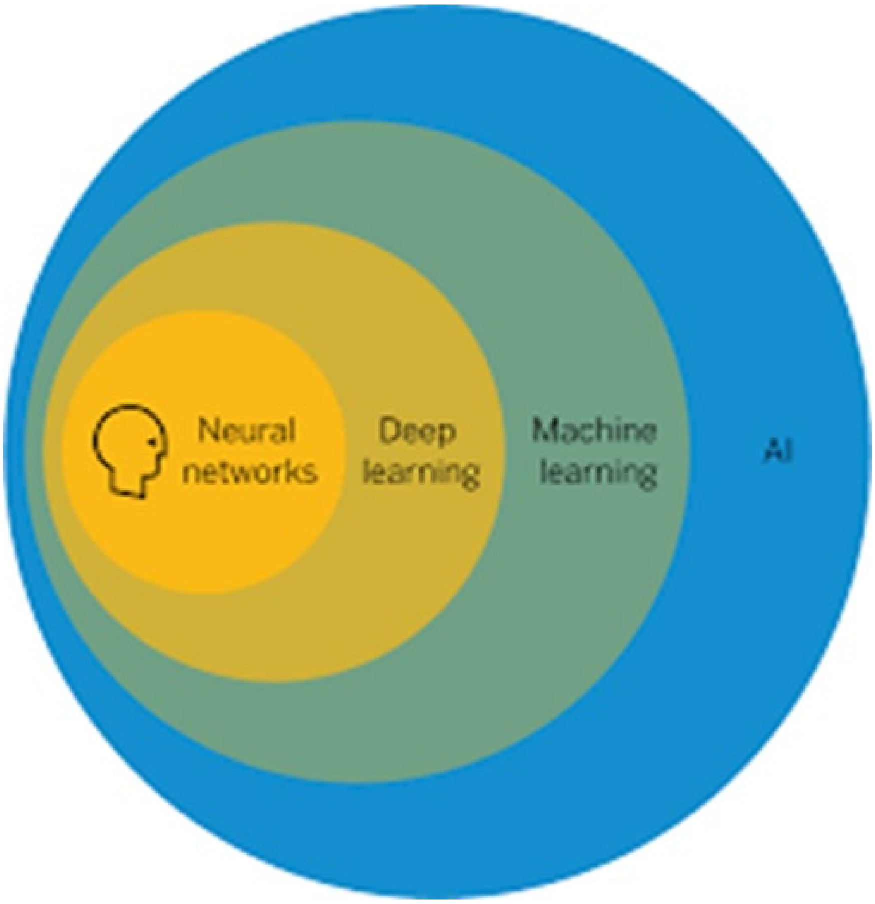
What New Data Research Repository Features Would Users Like to See? (Chan-Park, Sare and Waugh, 2022)

The final section of this research looks further at the 'analytical and computational' tools data feature-set request, focusing upon current data-centered researchers. Two examples are utilized from Deep Learning AI and Neural Net computational research, highlighting how the datasets from these areas are usefully being globally enabled and shared through present data research repositories and ecosystems.[4]

[3] See Uzwyshyn, 2021. Frameworks for Long Term Digital Preservation.

[4] Beyond the scope of this article but towards the need for 'algorithmic' literacy of researchers outside of Computer Science disciplines is enabling the vast potential of 'other' disciplinary area datasets towards the potential of insight and discovery possible through AI and Scholarly communication possibilities of libraries.

5 DATA RESEARCH REPOSITORIES, DIGITAL ECOSYSTEMS AND AI



Relationships among AI and Subdomains of Machine Learning, Deep Learning and Neural Nets

The last five years (2017-2022) have shown incredible progress and gains in analytical computations tools and discovery, particularly those tools and methodologies associated with new domains of Artificial Intelligence and these areas. Machine learning, deep learning and neural net scientific research has shown incredible scientific breakthroughs. These breakthroughs range from fields of Computer Vision (Facial/Object Recognition), Natural Language Processing (speech to text recognition and translation), Cybersecurity and Fraud Detections, Conversational Chatbots and Robotic Agents and Strategic Reasoning (AlphaGo, Game Theory). Scientific breakthroughs here have been enabled through a fortuitous combination of better algorithms plus greater computing processing power (Compute) plus notably readily available well labelled and metadata enabled online datasets, through increasingly open-source research data repositories and ecosystems.

The following section utilizes recent discoveries from Neural Net object identification to illustrate how online data research repositories and online data research ecosystems are facilitating the next generation of global collaboration possible with networked ecosystems academic research, discovery, and open science possibilities.

6 CANCER DETECTION, IMAGE DATA REPOSITORIES & AI

In 2017, an innovative new cancer detection methodology was published in Nature by a Stanford University group proposing the use of Neural Nets (Esteva, Nature, 2017). The AI neural network was trained on big data and dataset of 129,460 images of 2,032 diseases and larger dataset AI training images (1.41 million) to classify skin cancer lesions with deep neural networks. After comparison, the neural net machine learning AI did equal to or better than 30 board certified dermatologists with decades of experience.

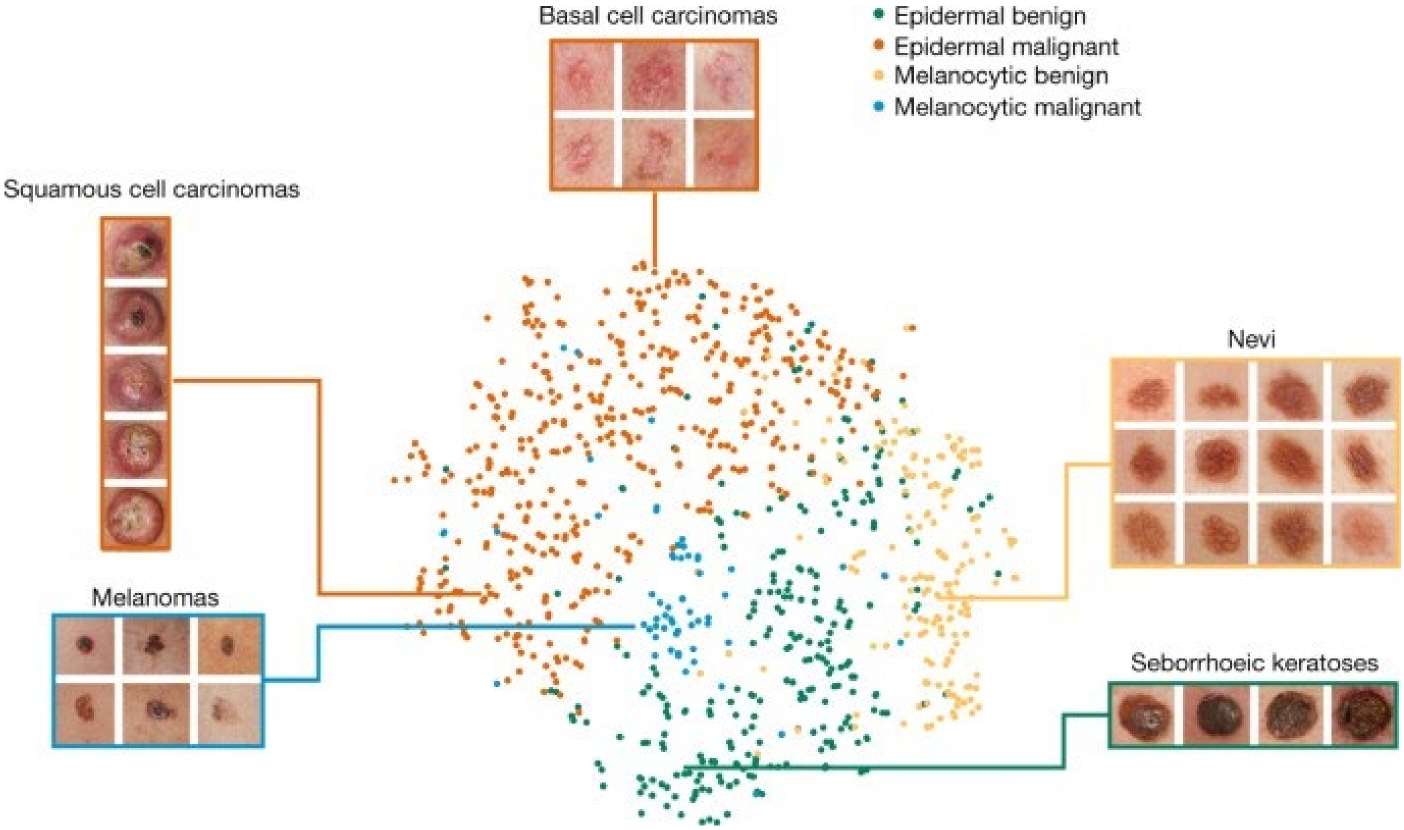


Image from Dermatologist Level Classification of Skin Cancer with Deep Neural Nets (Esteva et al, 2017) [5]

The neural net here was able to successfully classify epidermal lesions for early cancer detection into benign and cancerous (malignant) lesions better than 30-year board certified dermatologists by going down to pixel level differentiation levels and being trained through a multi-level neural net AI model. The large relevance of the digital image data repositories for initial training and metadata labelling should not be underestimated for researchers or significance of these methodologies. In a more recent article on Deep Learning in Cancer Pathology Surrounding a New Generation of Clinical Biomarker (Echle, 2020), the authors similarly emphasize both the need for organized digital libraries and data repositories and digital datastet preparation and metadata preprocessing for later accuracy in training, testing, and external neural net validation. The next example builds on the Stanford new discovery and possibility through possibilities now available through data repositories and digital scholarly ecosystems.

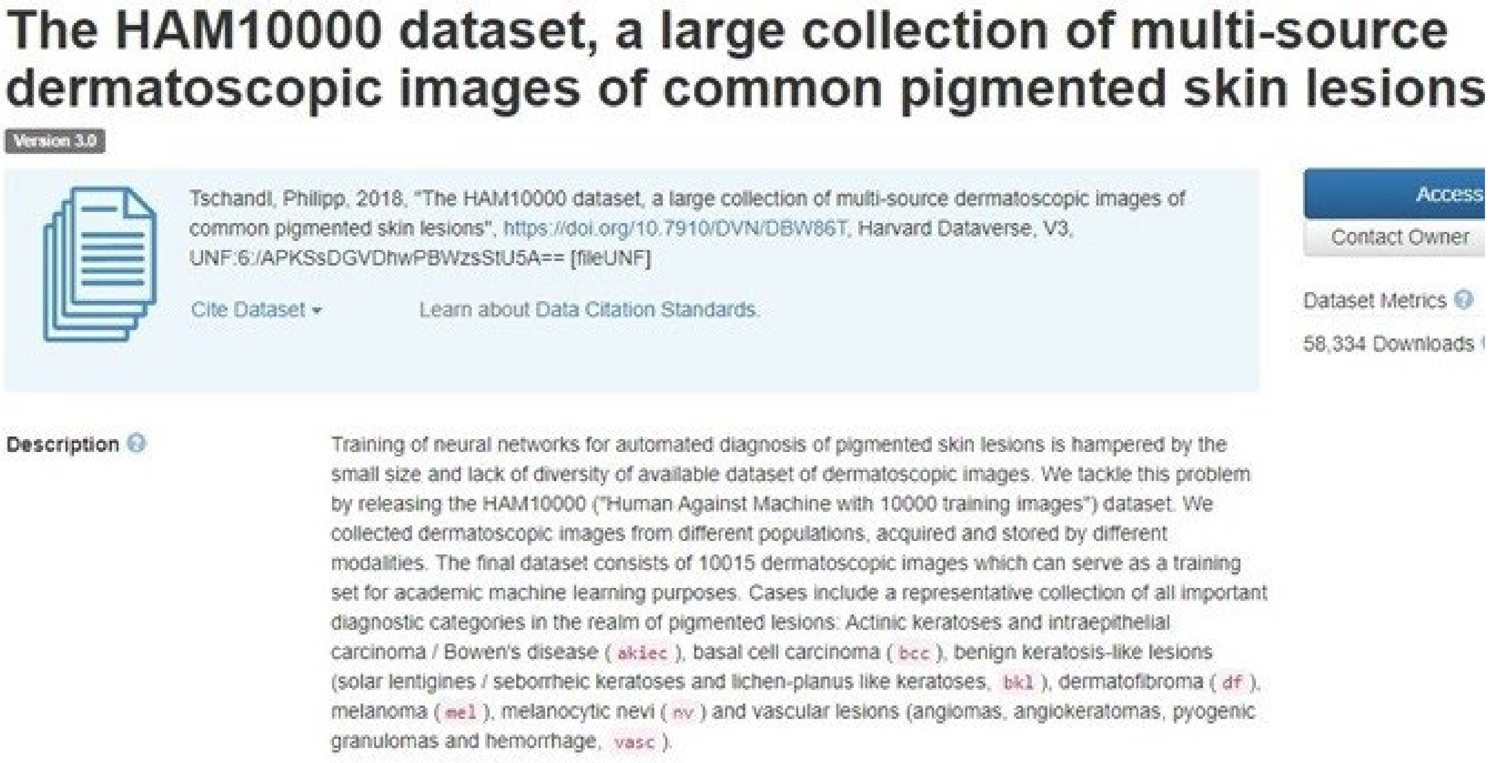
[5] See also, the original article from Nature. Esteva, A, Thrun, S. et al. Dermatologist-level Classification of Skin Cancer with Deep Neural Networks. Nature, Volume 542 (February 2, 2017). pp. 115-119. doi:10.1038/nature21056

7 OPEN SCIENCE, DATA, AI AND DIGITAL SCHOLARLY ECOSYSTEMS

Huge data sets like the Stanford example above are not the only and most recent of those able to be utlized through AI and neural net methodologies. Innovative global open science and AI machine learning possibilities are now being forwarded more efficiently through previous algorithmic training and application of regular sized datasets. New affordances are enabled through a confluence of data research repositories presently online, researchers' willingness to share their data sets online, and research data libraries open search and retrieval policies allowing other researchers willingness to apply their algorithmic machine learning expertise to research data.

Because Harvard's Dataverse also allows for the uploading of datasets from other universities globally, appropriate research datasets may be uploaded for sharing later or use by any researchers globally. If a university or research institution does not possess a Texas Data Repository or Harvard Data repository, and the researcher is carrying out valid academic research, they can utilize the Harvard repository. As mentioned though, Dataverse is open source software and any research level libraries, institutions and universities should be enouraged to be setting up their own instances of both data repositories and digital ecosystems.

To trace a current innovative example, the HAM10,000 image dataset below is a large collection of multi-source dermatoscopic images of cancerous skin lesions uploaded to Dataverse by Viennesse Dermatologist, Dr. Philip Tschandl, in 2018, a year after the Stanford Nature Neural Net algorithmic methodology article appeared.



The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions

Version 3.0

Tschandl, Philipp, 2018, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions", https://doi.org/10.7910/DVN/DBW86T, Harvard Dataverse, V3, UNF:6:/APKSsDGVDhwPBWzsStU5A== [fileUNF]

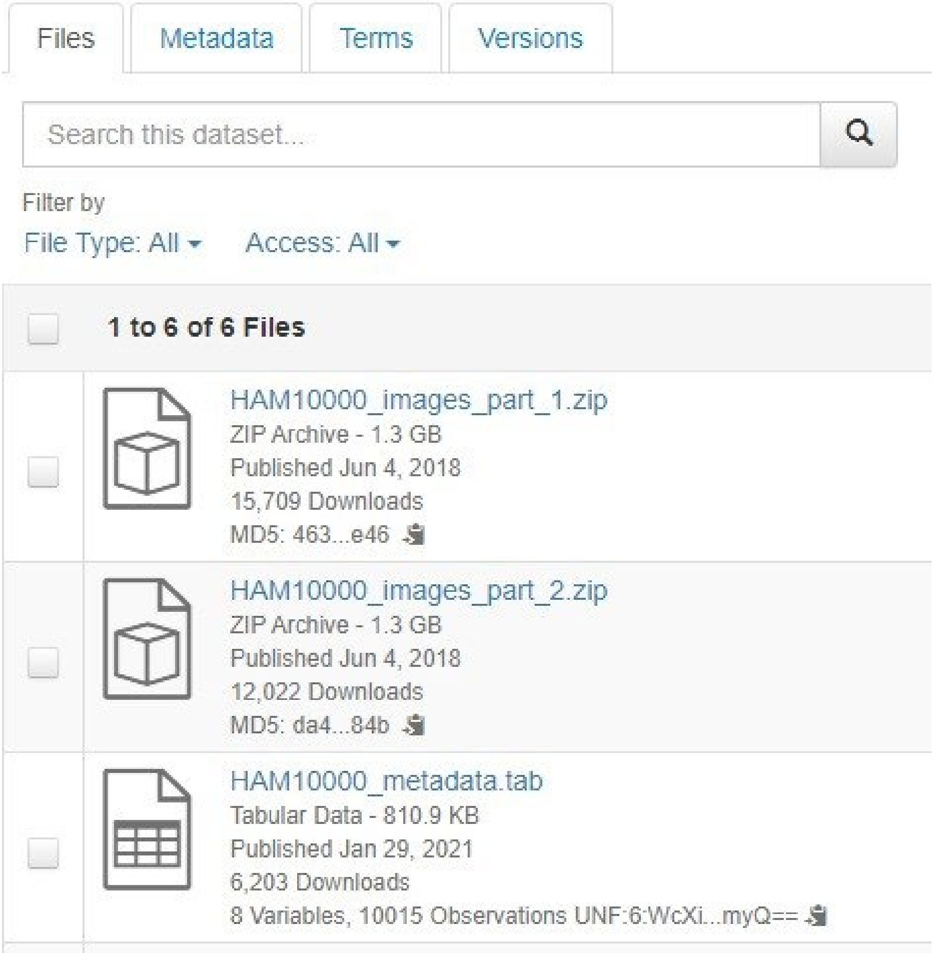
58,334 Downloads

Description

Training of neural networks for automated diagnosis of pigmented skin lesions is hampered by the small size and lack of diversity of available dataset of dermatoscopic images. We tackle this problem by releasing the HAM10000 ("Human Against Machine with 10000 training images") dataset. We collected dermatoscopic images from different populations, acquired and stored by different modalities. The final dataset consists of 10015 dermatoscopic images which can serve as a training set for academic machine learning purposes. Cases include a representative collection of all important diagnostic categories in the realm of pigmented lesions: Actinic keratoses and intraepithelial carcinoma / Bowen's disease (`akiec`), basal cell carcinoma (`bcc`), benign keratosis-like lesions (solar lentigines / seborrheic keratoses and lichen-planus like keratoses, `bkl`), dermatofibroma (`df`), melanoma (`mel`), melanocytic nevi (`nv`) and vascular lesions (angiomas, angiokeratomas, pyogenic granulomas and hemorrhage, `vasc`).

HAM10000 Dataset in Dataverse Data Research Repository,
https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T

As can be seen below, the images and metadata can be easily downloaded, unzipped, and used by researchers for neural net training purposes.



□ 1 to 6 of 6 Files

□ HAM10000_images_part_1.zip
ZIP Archive - 1.3 GB
Published Jun 4, 2018
15,709 Downloads
MD5: 463...e46

□ HAM10000_images_part_2.zip
ZIP Archive - 1.3 GB
Published Jun 4, 2018
12,022 Downloads
MD5: da4...84b

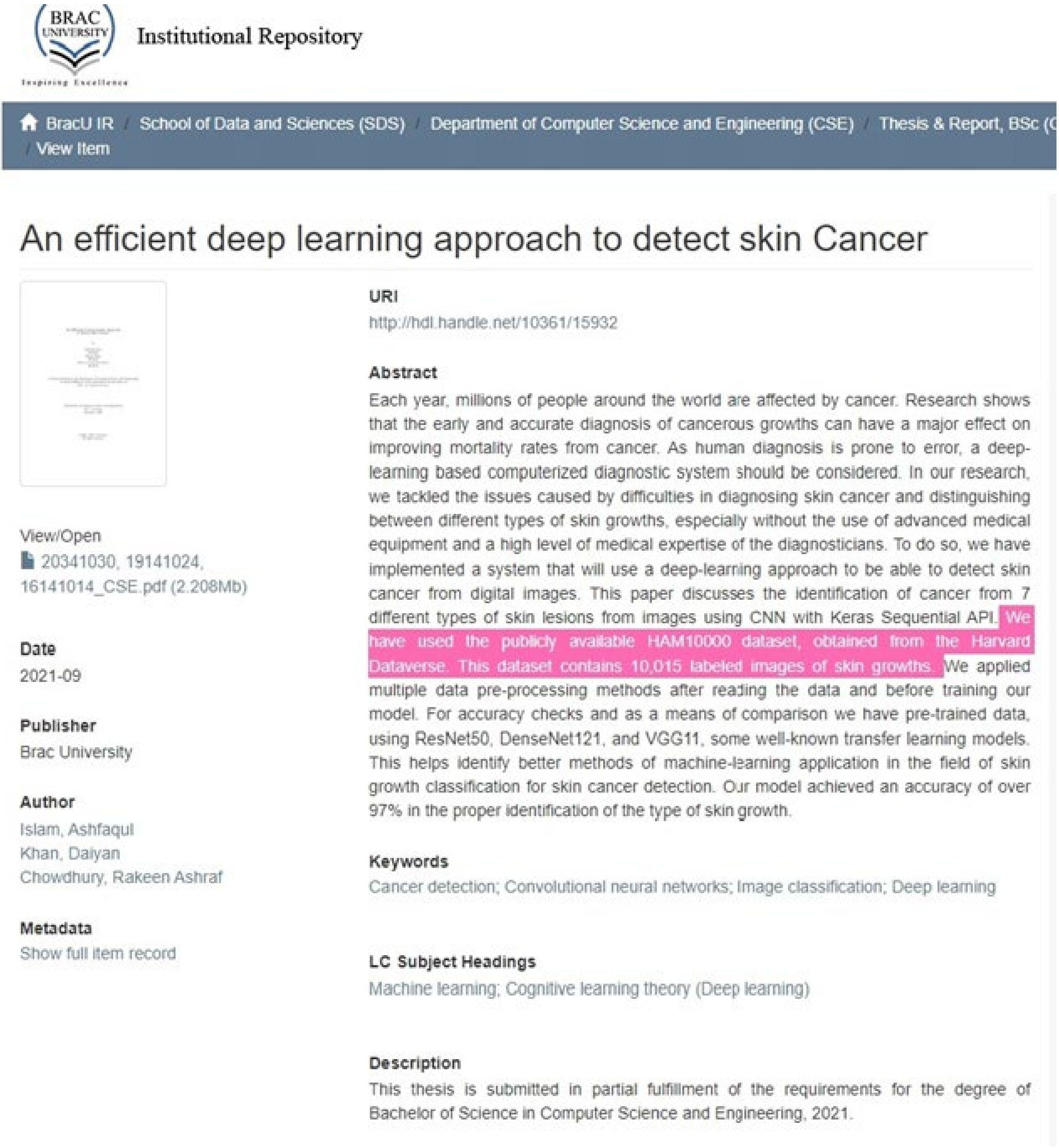
□ HAM10000_metadata.tab
Tabular Data - 810.9 KB
Published Jan 29, 2021
6,203 Downloads
8 Variables, 10015 Observations UNF:6:WcXi...myQ==

Dermascopic Cancer Images, Harvard Dataverse Repository,
https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T

The use of data research repositories to house data collections from around the globe, and later reuse by other researchers in other areas of the globe to train deep learning and neural net models, becomes very interesting with regards to possibilities for open science, globally dispersed academic researchers, and new possibilities for discovery and innovation.

Below is a cover page from BRAC University from Dhaka Bangladesh that uses DSpace as an institutional repository to house theses and dissertations from the School of Data and Sciences, Dept. of Computer Science and Engineering. Here, the computer science and engineering students had earlier downloaded Dr. Tschandl’s uploaded dermatological cancer training images, metadata and datasets and utilized the labelled image data as training material to train a deep learning neural net algorithm to recognize cancer growths with efficiency greater than, or equal to the 2017 board certified dermatologists for mobile devices. The example is very interesting for both possibilities of telemedicine and global populations which may not have as quick access to highly trained specialists as those in the West.

This is a particularly good example of open science and AI possibilities operating on global levels through the enabling power of digital scholarship ecosystems and data repositories. Content and specialized image data sets, with highly specialized labelled metadata that otherwise would be unavailable, are brought together with new machine learning algorithmic techniques and new research and an exceptionally good thesis has been produced. Globally dispersed content and data, from three different continents, has been aggregated to advance the pursuit of knowledge and science with a speed and utility that would be unimaginable in other centuries.



BRAC UNIVERSITY Inspiring Excellence Institutional Repository

BracU IR / School of Data and Sciences (SDS) / Department of Computer Science and Engineering (CSE) / Thesis & Report, BSc (
/ View Item

An efficient deep learning approach to detect skin Cancer

View/Open
20341030, 19141024, 16141014_CSE.pdf (2.208Mb)

Date
2021-09

Publisher
Brac University

Author
Islam, Ashfaqul
Khan, Daiyan
Chowdhury, Rakeen Ashraf

Metadata
Show full item record

URI
http://hdl.handle.net/10361/15932

Abstract
Each year, millions of people around the world are affected by cancer. Research shows that the early and accurate diagnosis of cancerous growths can have a major effect on improving mortality rates from cancer. As human diagnosis is prone to error, a deep-learning based computerized diagnostic system should be considered. In our research, we tackled the issues caused by difficulties in diagnosing skin cancer and distinguishing between different types of skin growths, especially without the use of advanced medical equipment and a high level of medical expertise of the diagnosticians. To do so, we have implemented a system that will use a deep-learning approach to be able to detect skin cancer from digital images. This paper discusses the identification of cancer from 7 different types of skin lesions from images using CNN with Keras Sequential API. We have used the publicly available HAM10000 dataset, obtained from the Harvard Dataverse. This dataset contains 10,015 labeled images of skin growths. We applied multiple data pre-processing methods after reading the data and before training our model. For accuracy checks and as a means of comparison we have pre-trained data, using ResNet50, DenseNet121, and VGG11, some well-known transfer learning models. This helps identify better methods of machine-learning application in the field of skin growth classification for skin cancer detection. Our model achieved an accuracy of over 97% in the proper identification of the type of skin growth.

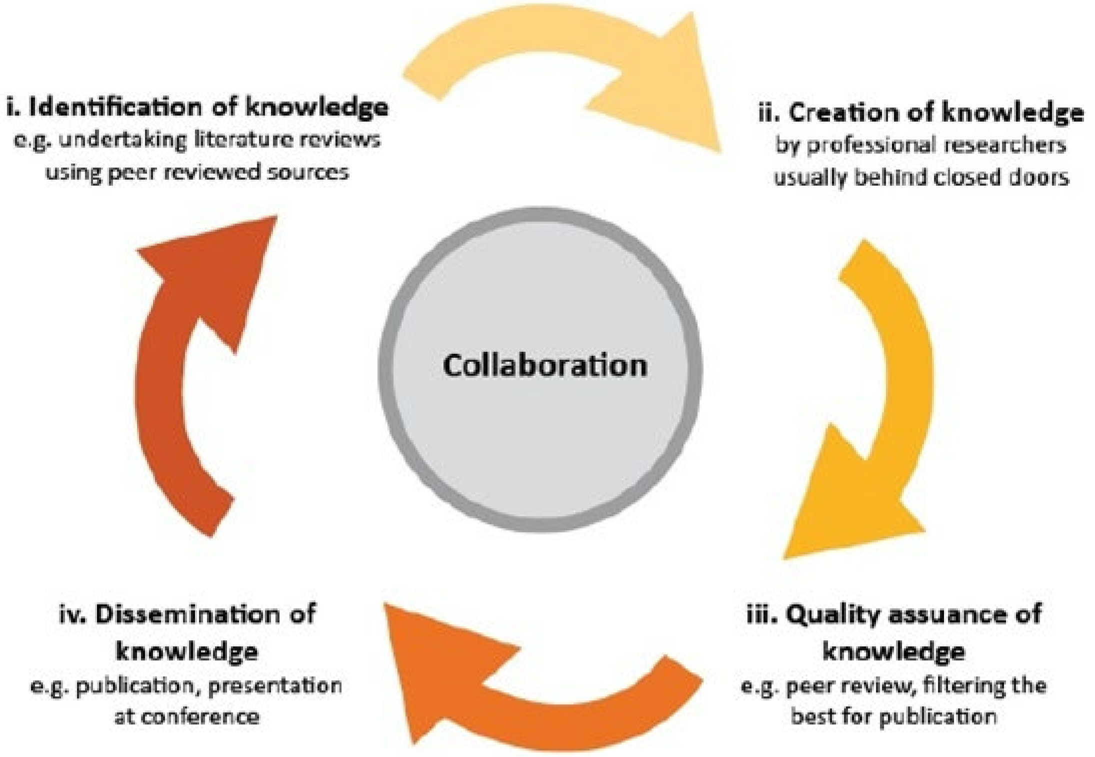
Keywords
Cancer detection; Convolutional neural networks; Image classification; Deep learning

LC Subject Headings
Machine learning; Cognitive learning theory (Deep learning)

Description
This thesis is submitted in partial fulfillment of the requirements for the degree of Bachelor of Science in Computer Science and Engineering, 2021.

BRAC University Dspace Repository 2021 Deep Learning/AI Thesis http://dspace.bracu.ac.bd/xmlui/handle/10361/15932

8 CONCLUSIONS – AI, DATA AND THE ACADEMIC RESEARCH CYCLE


i. Identification of knowledge
e.g. undertaking literature reviews
using peer reviewed sources
ii. Creation of knowledge
by professional researchers
usually behind closed doors
Collaboration
iii. Quality assuance of
knowledge
e.g. peer review, filtering the
best for publication
iv. Dissemination of
knowledge
e.g. publication, presentation
at conference

The Academic Research Cycle, Cann, Dimitrious and Hooley, 2011.

New data repository and digital scholarly ecosystem possibilities are enabling the academic research cycle and progress of knowledge and discovery in our new millennia in amazing ways. Through the enabling power of data, data research repositories and digital scholarly ecosystems, open science possibilities enable a new globally networked generation towards incredible new science and knowledge discovery and creation.

The creation of data and knowledge usually occurs behind closed doors, hidden away in research labs, file cabinets and, more recently, computer hard drives. Data sharing has now been enabled through possibilities of networked communication and content technologies. This sharing by researchers on a global stage also allows transparency towards the quality assurance of knowledge ranging from online peer review to availability of data and research for further citation, discovery download and pragmatic use.

Paired with other ecosystem possibilities such as open online academic research journals, theses and dissertation (VIREO) and online identity management systems (ORCID), and new multimedia user interface possibilities, these tools are able to facilitate large global collaboration and intrinsically human creative activities of discovery and invention, creating new innovation and building on the progress of previous and current generations of researchers and scholars.

REFERENCES

Artificial Intelligence. Machine Learning. Neural Networks. Future Technology. Bloomberg Businessweek Canada. 2022. https://www.youtube.com/watch?v=ypVHymY715M

Cann, A., Dimitriou, K. Hooley, T. Social Media: A Guide for Researchers. Research Information Network. University of Derby, UK, 2011.

Chan- Park, C. and Sare, L. Waugh, S. *Results of the Texas Data Repository User Survey*, 2022. Texas Conference on Digital Libraries Presentation, 2022.

ColdFusion (2018). *Why Deep Learning Now?* (Documentary Overview). https://www.youtube.com/watch?v=b3IyDNB_ciI

Echle et al. Deep Learning in Cancer Pathology: A New Generation of Clinical Biomarkers. *British Journal of Cancer*. November 2020. https://www.nature.com/articles/s41416-020-01122-x

Esteva, A, Thrun, S. et al. Dermatologist-level Classification of Skin Cancer with Deep Neural Networks. *Nature*, Volume 542 (February 2, 2017). pp. 115-119. doi:10.1038/nature21056

Fridman, Lev. *MIT Deep Learning and Artificial Intelligence Lectures*. https://deeplearning.mit.edu/ 2022.

Islam, A., Khan, D. and Chowdhury, R. 2021. An Efficient Deep Learning Approach to Detect Skin Cancer Undergraduate Thesis. BRAC University DSpace Institutional Repository, 2021. Available: http://dspace.bracu.ac.bd/xmlui/handle/10361/15932

Mitchell, Tom. 2022 *Where on Earth is AI Headed?* Carnegie Mellon. https://www.youtube.com/watch?v=ij9vqTb8Rjc

Peters, T. and Waugh, L. Larger Data Storage Report: Research Data Management Initiatives and Planning, January 2022. Texas State University Libraries (Unpublished White Paper)

Texas Data Repository 2022. https://dataverse.tdl.org/

Tschandl, Phillip et al. Human-computer Collaboration for Skin Cancer Recognition. Nature Medicine, 22 June 2020, 1229-1234. See: https://www.nature.com/articles/s41591-020-0942-0

Uzwyshyn, R. 2020. *Developing an Open-Source Digital Scholarship Ecosystem*. International Conference on Education and Information Technology ICEIT2020. St. Anne's Oxford, United Kingdom. February 2020. Available at: https://www.researchgate.net/publication/336923249_Developing_an_Open_Source_Digital_Scholarship_Ecosystem.

Uzwyshyn, R. 2021. Frameworks for Long Term Digital Preservation Infrastructures. *Computers in Libraries*. September 2021. pp.4-8.

- - -. *Open Digital Research Ecosystems: How to Build Them and Why*. Computers in Libraries, (40) 8. November 2020. https://www.researchgate.net/publication/345956074_Online_Digital_Research_Ecosystems_How_to_Build_Them_and_Why

---. Online Research Data Repositories: The What, When Why and How. Computers in Libraries. 36:3, April 2016. pp. 18-21. http://rayuzwyshyn.net/TXU2016/OnlineDataResearchRepositoriesUzwyshyn.pdf

Waugh, L. *Texas State University Annual Usage Report 2020.* TXST Dataverse Repository. Texas Conference on Digital Libraries Presentation. Texas State University.