## Data to Science Engine (D2SE) A Data-Driven Open Science Ecosystem for Sustained Innovation

by Jinha Jung Associate Professor @ Lyles School of Civil Engineering, Purdue University



## 10

#### The Greatest Inventions

## In the past 1,000 years



Lyles School of Civil Engineering

Source: https://ehistory.osu.edu/articles/greatest-inventions-past-1000-years





#### 8. Airplane (1903)













#### 2. Electric Light (1879)



## 1. Printing Press (1450)

Inneit spillela fandt throntini ab and man professore of consider tot ar hydror libro-capitali post. 6 arr aubzofue na wichi mutufnia pizzo-fruite af a fantisticae tra fa primaje

amininar-fitt pheet iam libai a ummis aminine noma: plenbaur. Hmamidia uterfireado i-- ונשט מסמ מינה הזה נעמי סממולה וקר ו tao m fanufiano-nő púna cancen mepte-mi litola a palpãe abulariofit bei einen- er binnnan feriprurarti fludia conciliant, lenon? in ucenibs hiltorio-quoloa luftraffe puinciaf. nonce aprille mine-mana mililleut cos pucos or libris unumaur; casa is vident, Benir piragoan momphinitus unito-fir place origini-a auchini carounium-gautomes geam vealur-out quoubă magua puoia binbaitlatorichillour preasure or ur qui achenio unfr mar-sporme-minifes bodyman advatnuir gignalia plonabär-hom pojnus any bilioplus-malés stima motér biliopriju lua iputori ingori. Durin ni leas quali con acht higimno pliquit capit a picado a umilhame orino midilimo panir bud? capnung wind? a frung, Tami auia pling mains pupper is hir-ab press linin ladeo doquéer fonter manantéte ultimite hilpenie galliarien huits-quolitari umille nobileo leginuus-a quos ab secuplacione fui roma no maexacione bois fama phueir. Pabuir illa man inaubini onihi feculiaciducandoire miradin-ur urbi ranni

-ix

intereffication green actions queeneer. Ruolloni" fiur ille man? ur ouleuf loquiner-fine plus-ur piragona nabunt-inmanic plas-pailinir mundualbanos fridas-malfarmas-opulinilinna indig rogna provenanie ir ab errouten faciliano photon ampur elifiuille puenir ab bragmanas er hranam in throng from arrow it to cansali foner pocament-iner panco" bifeipiosche narura-te monto- ar te cufu bim et fibre aubier boenen. Autr p daminao-babilonipo-dalle. pe-michos-affricos-particos-fricos-pleminos-acabro-palefinicos-milus ab allmanbria:promir ab obiopia-ur gignolophiBao a famolifim am folis menfam niberr in fabuto. Anutuit ille uit ubits to bifente a fortp profinito-famp fe mitioe finm. Sena. fir fuper hor planiffinge ada voluminibus:phylofinanio.

uib loguar br fedi bominibsni aplice paulue cuas decioni. a magifter genoù-qui te confirma nin i le holpino loquebar biero. An provintation quantino dino qui in sur loquit pir, polt mmalin arabidirs fuffcasi: afeithir ihenfoliuna ur uitir paren a main aputo ni birbo gumbrei. for mi mifio robomabio r ogoo abischung genni übirace inftrumbus ctar. Butfur polt anos getortecim affiniupio barnaba er procego foir ni aphs coandin-ne fore in vamum monor aur meuroller. haber nefeio qu laennio energie: vine voor adus er itt aures bilapli te autoris pur mauffula:formus louar, diutr or nichingus qu'entri mularer-s legender







#### **Mass Media**



### Independent Media





• <b>dvd</b> Search term	• youtube Search term	+ Add comparison
Worldwide	All categories 💌 Web Search 💌	
Interest over time ⑦		⊻ <> ≪
100		
75		
50		
25		
50 25 Average Mar 1, 2004	Nov 1, 2006	Jul 1, 2009







Indian record label T-Series is the mostsubscribed channel, with over 263 million subscribers as of April 2024. American YouTube personality MrBeast is the most-subscribed individual and second mostsubscribed channel overall, with 252 million subscribers as of April 2024.

263 M Subscriber

https://6sense.com/tech/media-players-and-streaming-platforms/youtube-market-share

#### **Traditional Remote Sensing Platforms**



5/3/24 18

#### Modern Remote Sensing Platforms







## Building an open-source online platform

**One-stop shop researchers can use to...** 

Manage	Process	Visualize	Collaborate	Publish
Organize uploaded datasets and make accessing data easy from the D2S platform.	Convert datasets to cloud optimized formats and provide tools for deriving new data products.	Visualize raster and point cloud datasets on interactive maps and viewers.	Create teams and manage access to projects with simple controls. Share findings with other users on the platform.	Publish data products to a public catalog for anyone to discover.



#### Cloud Optimized File Formats

- Automatically convert uploaded GeoTIFFs to Cloud Optimized GeoTIFF (COG) format and LAS point clouds to Cloud Optimized Point Cloud (COPC) format
- GDAL used to create COG in asynchronous background task
  - Prevents blocking user from interacting with site while waiting for conversion to complete
- Untwine used to create COPC in background
- Stream only required portions of COG and COPC data instead of downloading entire dataset





### Visualizing COGs on D2S

- Dynamic tile server, TiTiler, accesses COG stream and generates map tiles
- Leaflet's TileLayer seamlessly interacts with dynamic and static map tiles
- TiTiler generates new map tiles on-the-fly when symbology properties updated in UI
- Successfully used on datasets exceeding 50 GB





Screenshot of a Cloud Optimized GeoTiff in D2S.



### Visualizing COPCs with Potree on D2S



June 4<sup>th</sup>, 2022 - RGB



June 23<sup>rd</sup>, 2022 - RGB



July 25<sup>th</sup>, 2022 - RGB



June 4<sup>th</sup>, 2022 - Elevation



June 23<sup>rd</sup>, 2022 - Elevation



July 25<sup>th</sup>, 2022 - Elevation



## Share maps

Share maps with the public or restricted to only D2S users with appropriate permissions

⊧ HOM				
		×		
25, 2022			and a summarized state of the second se	
) Sensor: R				
0 n 23, 202	Share Se	ttings		
) Ser O	Restricted		Only project members will be able to access shared links for this data product. Must be signed in to platform to view shared link.	enemente coltragato contragato contrato
	Anyone		Anyone can access this data product. It can be downloaded, used outside of the platform, and shared links can be viewed without signing in.	
) Max: C	Copy file link	k 🛛 Inclu	de symbology	nennen nennen nennen nennen
tings	~		-0.2 -0.4 C Refresh	
Data	products of	can be s	set to restricted access or anyone (public).	



Maps for publicly shared data products can be viewed by anyone.



## Accessing your data

Data stored on D2S can be accessed from external applications that support COG and COPC formats











## Geospati al Data

### https://docs.gdsl.org/data-to-science-engine-workshop/



AG2P1

Agricultural Genome to

Phenome Initiative



**College of Agriculture** 



### Acknowledgements

• This research was supported [in part] by the intramural research program of the U.S. Department of Agriculture, National Institute of Food and Agriculture, Grant (USDA-NIFA award 2022-70412-38454).

The findings and conclusions in this preliminary presentation have not been formally disseminated by the U.S. Department of Agriculture and Should not be construed to represent any agency determination or policy.

- Purdue Plant Sciences 2.0 Initiative
- Purdue Digital Forestry



Phenome Initiative



**College of Agriculture** 



# Thank You

#### **Development Team Contact**

Jinha Jung (jinha@purdue.edu) Project Lead

Minyoung Jung (jung411@purdue.edu) Project Manager

Cheryl Qian UI/UX Lead



Ben Hancock (hancocb@purdue.edu) Web Developer

Na Zhou UI/UX Designer