

Cyber Tools for Research: a Tour

Ouida W. Meier, Ph.D.

Program Manager, EarthCube CRESCYNT – Coral Reef Science and
Cyberinfrastructure Network, University of Hawaii

omeier@hawaii.edu or crescyntcn@gmail.com

last updated 2016-04-29

Tools for:

- Facilitating teamwork
- Planning and executing research
- Cleaning data
- Accessing and incorporating standards
- Storing data
- Analyzing data
- Communicating results
- Making data and results accessible

Bottom line for EarthCube

To massively increase functional capacity in geosciences (broadly defined) by taking advantage of advanced computing capabilities, especially through:

- Data interoperability
- Data discovery
- Data reuse

(including many different types of data)

...in order to answer the **highest priority science drivers and research questions**, and address issues of highest priority for society.

Disclaimer and Invitation

- Opinions expressed are my own, not EarthCube's, NSF's or CRESCYNT's
- If you like and use other tools in any of these categories, please post their names or links in the chat box and I'll add them to the list
- If you would like a direct copy of this list, please revisit the science tools webinar series webpage later or send your email address (to presenter only) in the chat box
- Please list or describe your other tools needs in the chat box as well!
- Please consider joining EarthCube to continue making collective progress!
- There are separate listings for datasets; this listing is for data tools.

Collaboration – video, teleconferencing, desktop sharing

- [Skype](#)
- [Google Hangout](#)
- [WebEx](#) \$
- [GoToMeeting](#) –Citrix \$
- [Adobe Connect](#) \$
- ~~ooVoo~~ – do not do this to yourself

Collaboration – shared workspace, collective products

- [OSF Open Science Framework](#)
- [Authorea](#) (collaborative writing including data and code)
- [Overleaf](#) (collaborative writing and publishing)
- [ShareLaTeX](#) (online group LaTeX editor)
- [Slack](#) (communication, accepts [add-ons](#) – e.g., [Asana](#) for task scheduling, [Drive](#), etc.)
- [Trellis](#) (initiative of AAAS, currently in beta development)
- [Basecamp](#)
- [Google Drive](#) (shared folders, docs, sheets)
- [Dropbox](#)

[Alternative collaborative workspaces](#), [shared-drive comparisons](#)...

Settings

[Profile information](#)

[Account settings](#)


[Configure add-on accounts](#)

[Notifications](#)

[Developer apps](#)

[Personal access tokens](#)

Configure Add-on Accounts

 Amazon S3

Access Key

Secret Key

Save

 Box

[Connect Account](#)

 Dataverse

[Connect Account](#)

 Dropbox


[Connect Account](#)

 figshare

[Connect Account](#)

 GitHub

[Connect Account](#)

 Google Drive

[Connect Account](#)

 Mendeley

[Connect Account](#)

 Zotero

[Connect Account](#)

Workflows

- [yEd](#) *
- [gliffy](#)
- [Microsoft PowerPoint](#), [Visio](#)
- [OpenOffice Draw](#), [Impress](#); [WPS Office](#); [Zoho Docs](#)

*Potential expansion capacities to consider:

adapting layouts for different purposes and audiences

machine-readable ([GraphML](#)) – both text and visual layout

potential use as discovery tool (e.g., iPlant now [CyVerse](#))

- [The Kepler Project](#) – integrated scientific workflow system

Community-Based Platforms

- [CyVerse](#) – was [iPlant](#) – NSF data management platform
- [NIF Neuroscience Information Framework](#)

- [Agave](#) (NSF sponsored, free to use, developed by TACC, e.g., Agave API powers CyVerse, science-as-a-service, very powerful) - [tutorials](#)

Repositories

- Whatever discipline-specific repository your field uses as consensus
- Publication-specific or project-specific requirements
- [Zenodo](#) (open section of CERN)
- [Dataverse](#) (Harvard)
- [figshare](#)
- [Dryad](#)

Pointer: [DataCite](#)

Clearinghouse: [DataONE](#)

Statistical Data Analysis

- [R](#) (many packages)
- [Python](#) (many libraries)
- [SAS](#) (and [JMP](#))
- [SPSS](#) (bought by IBM)
- [PSPP](#) (freeware version)
- [Statistica](#) (bought by Dell)
- [Systat](#) (and [SigmaPlot](#))

Exploratory Data Analysis

- [Rapid Miner](#)
- [Weka](#)
- [R](#), [Python](#), [SAS](#) & [JMP](#)...

Visual Exploration:

- [Ocean Data View](#)

Data Visualization Tools

- [R](#), [RStudio](#), [Shiny](#) (interactive web application)
- [ggplot2](#) – on top of R; Grammar of Graphics
- [Watson Analytics](#) (IBM; was Many Eyes)
- [Tableau](#) and [Tableau Public](#)
- [Excel](#) (incl. PowerPivot graphs)
- [Gephi](#)
- [plotly](#)
- [MATLAB](#)
- javascript: [D3.js](#), [Processing](#), [more](#) and [even more](#) js libraries
- Extra: lists of [5 best](#), [14 best](#), [38 best](#) data visualization tools

Mapping – Spatial Visualization & Analysis

- [ArcGIS Online](#)
- [ESRI ArcGIS](#)
- [Google Earth](#) (being replaced by [Earth Engine](#))
- [ggplot2](#) (R package)
- [Maptive](#)

Storing and Accessing Research Products

Repositories for Data and Metadata

Analysis Products

Identifiers

Publication, Reuse

Protocols

- [Scientific Protocols](#) – built on GitHub
 - [protocols.io](#) - versioning
 - [protocol exchange](#) - Nature
 - [Protocol Online](#)
 - and in repositories: [figshare](#), [Zenodo](#), etc.
-
- [VisTrails](#)

Unique Identifiers

- [DOI](#) to register/obtain a digital object identifier; [registration agencies](#)
- [IGSN](#) (at [SESAR](#)) for geosamples (and their subsamples)
- [ORCID](#) for individual researchers
- [CrossRef](#) for linking publications, [CrossRef API](#) for metadata search

Persistent Identifiers (less rigorous, self-maintained):

- [PURL](#) and guide to [PURLs in the Wild](#)

Metadata Creation

- [KNB Knowledge Network for Biocomplexity](#)
- [BCO-DMO](#)
- [Marine Metadata Interoperability Project](#)
- [EML Ecological Metadata Language](#) & [Morpho](#) (EML tool)
- [Metacat](#)
- [netCDF](#) Network Common Data Form

Standards

- [CSDMS](#) Community Surface Dynamics Modeling System – modeling and metadata
- [Digital Curation Centre](#) – list of metadata standards
- [Genomic Standards Consortium](#)
- [Geosemantics framework](#)

Image Management

- [Google Photos](#) (replaces Picasa after 2016-03-15)
- [Adobe Lightroom](#) \$ free trial
- [Flickr](#) – but as of 2016-03-09 its Auto-Uploadr only avail. in Pro version \$
- [PhotoBucket](#) – in practice mostly indiv. photographer distribution, sales

Metadata [e.g., Adobe Bridge (proprietary)] – microphotography

Analysis: [Image J](#) – open source, development funded by NIH

[Metamorph](#)

[Image-Pro](#) tools – MediaCybernetics \$

Sound Manipulation and Analysis

- [Audacity](#) –free and open-source sound recording and editing
- [Ocenaudio](#)
- [Wavosaur](#)

more free options at this [techradar review](#)

Bioacoustic analysis:

- [Song Scope](#) –meters and software, bats, land animals, marine animals \$
- [Raven](#) –Cornell; \$ and free lite version
- [Sonic Visualiser](#) –Queens Univ; graphical displays for visual analysis of sound

Biological Identification & Taxonomy

[GBIF Global Biodiversity Information Facility](#)

[ITIS Integrated Taxonomic Information System](#)

Discipline-Specific Auth., e.g., [WORMS World Registry of Marine Species](#)

Digitized collections, e.g., [iDigBio Integrated Digitized Biocollections](#)

Taxon search in databases, e.g., [KNB Knowledge Network for Biocomplexity](#)

CyVerse's [Taxonomic Name Resolution Service](#)

[BioLexicon](#) (bioinformatics terminology)

[Tree of Life](#)

Innovative education/citizen science tools:

[OneZoom](#)

[National Phenology Network](#)

[DiscoverLife](#)

[Zooniverse](#) (crowdsourced science assist)

[eBird](#)

[Galaxy Zoo](#) (astronomy, crowdsourced)

Data Exploration, Multiple Sources

- [Tableau](#) \$, acad. disc. if teaching or proposing course or enrolled in one
- [Tableau Public](#) –free online, plus public viewing for reg. Tableau projects
- [Microsoft Excel](#) pivot tables, dashboards underestimated; \$, acad. disc.
 - Many advanced Microsoft products avail to students free with Microsoft acct
- [Qlik](#) –free and \$ levels
- [R](#)
- [Python](#)
- **SEE ALSO PLATFORMS**

Learning to Code

- [CodeAcademy](#)
- [DataCamp](#)
- [Udacity](#)
- [Udemy](#)
- [Coursera](#)
- [edX](#) (free to audit, \$ certificate), [other MOOCs](#)
- [Hour of Code](#) (kids+)
- Specific-topic online tutorials (e.g., [SQLZoo](#), or [swirl](#) for R)
- [Software Carpentry](#)
- [Data Carpentry](#)

Reference, Web Citation, Annotation

- [Mendeley](#) – ref mgr and pdf organizer
- [Zotero](#) –devel by Mellon and Sloan Found's, Inst. Museum Libr. Svcs
- [Pundit](#) -European

Tool Descriptions, Reviews, Comparisons

- [Data Science Central](#)
- [KD Nuggets](#)
- [TechCrunch](#) (and [CrunchBase](#) – business)
- [Quora](#)
- [Wikipedia](#) (try: search, then browse by category)
 - e.g., “[Data mining and machine learning software](#)”

Desktop Screen Capture / Screencast / Video

- [OBS Open Broadcaster Software](#) – free & open source
- [Camtasia](#) (free trial)
- [Captivate](#) (Adobe)

Animation (with characters, backgrounds):

- [GoAnimate](#) (free trial)
- [OpenToonz](#) (2D - e.g., Studio Ghibli; released 3/27/2016, free & open source)
- [Unity](#) (up to 3D; free levels)

Lagniappe

- [ESIP Federation of Earth Science Information Partners](#) – open, networked community
 - [DIBBS Data Infrastructure Building Blocks](#) – NSF; including [Brown Dog Tool Catalog](#)
 - [VOEIS Virtual Observatory and Ecological Informatics System](#) – open source data mgmt
 - [NASA Global Change Master Directory](#) – datasets, services, tools, portals, collaboration
 - [ODI Open Data Institute](#) – global network, open data training courses
 - [RunMyCode](#) – repository for code testing (submit code and data – reproducibility, publ.)
-
- [UNEPLive](#) – United Nations Environment Programme – stored maps, graphs, publications
 - [Dato](#) – was GraphLab – machine learning platform, analytics output
 - [Alooma](#) - Amazon Redshift – integrated data services
 - [Azure](#) - Microsoft cloud computing platform & services
 - [CrowdFlower](#) – data enrichment platform

CREDITS:

- Yolanda Gil et al., EarthCube [OntoSoft](#), [Geoscience Papers of the Future](#)
- [DataONE list of tools and resources](#)
- Ilya Zaslavaky et al., construction of EarthCube [CINERGI](#) resource viewers

Feedback, Suggestions, Additions: Ouida Meier, crescyntrcn@gmail.com and [EarthCube CRESCYNT Coral Reef Science and Cyberinfrastructure Network \(http://crescynt.org\)](http://crescynt.org) – suggestions from first viewers have been added

Video and transcript will be posted at [EarthCube YouTube channel](#) and will be linked at the [EarthCube \(http://earthcube.org\)](http://earthcube.org) webinar series: [Doing Geoscience with EarthCube Tools](#)