



























Μεγάλα δεδομένα, νέα μέσα, νέοι τρόποι διαχείρισης, ανάλυσης και ερμηνείας: Επιτεύγματα και Προκλήσεις

Big Data, New Means new ways of management, analysis and interpretation: **Achievements and Challenges**



Project Coordinator

Ανοικτές Διαλέξεις/Συζητήσεις «Μεγάλα Δεδομένα, Νέα Μέσα, Ζητήματα Τεκμηρίωσης: Μαθαίνοντας από πρωτοπόρα εγχειρήματα»

Open talks/discussions "Big Data, New Means, Documentation Issues: Learning from pioneering attempts" Athens, NRF, March 20, 2019



GREECE









Competitiveness & Entrepreneurship

Executive Development Task Force

Chrysoula Bekiari bekiari@ics.forth.gr Martin Doerr martin@ics.forth.gr Nikos Minadakis minadakn@ics.forth.gr Nicolas Bailly nbailly@hcmr.gr Eva Chatzinikolaou evachatz@hcmr.gr Carlo Allocca carlo@hcmr.gr Tilemachos Bourtzis tbourtzis@gmail.com Sarah Faulwetter sarifa@hcmr.gr Niki Keklikoglou keklikoglou@hcmr.gr Jacques Lagnel lagnel@hcmr.gr Dimitra Mavraki dmavraki@hcmr.gr Matina Nikolopoulou snikolo@hcmr.gr Anastasis Oulas oulas@hcmr.gr Evangelos Pafilis pafilis@hcmr.gr Emmanouela Panteri emmipan@hcmr.gr Nikolas Pattakos pattakosn@hcmr.gr Christina Pavloudi cpavloud@hcmr.gr Theodore Patkos patkos@ics.forth.gr Kostas Varsos varsosk@ics.forth.gr





National Strategic Reference Framework and Technology - Hellas



Outline

- (a) Introduction
- (b)The MedOBIS Vitrual Laboratory (vLab)
- (c) The micro-CT vLab
- (d) The R vLab



tch The field: biodiversity

Definition(s)

..."the variability among living organisms from all sources including inter-alia, terrestrial, marine and other aquatic ecosystems and the ecological complexes of which they are part" (CBD)

...."includes diversity within species between species and ecosystems"..... (Gaston & Spicer, 2004)



LifeWatch Biodiversity is...

Genes and DNA

106 to 109 nucleotides in a DNA molecule



Species (organisms and their populations)

>10⁷ species; each species with 10² - 10¹² individuals



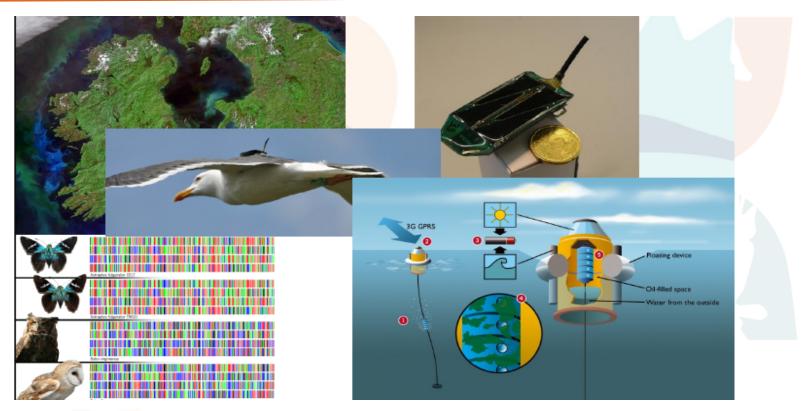
Ecosystems

habitats with 104 to 106 species, and manifold interactions





LifeWatch: the "data era"





Scientific approach

— Patterns

— Processes

— Consequences from changes



What do we need for its study?

Data

- + Observatories
- + Capacity
- + Networks
- + People
- -----
- = Infrastructure



The concept of ESFRIs: LifeWatch





ESFRIs: a new era of megascience

- Computational <u>capacity free</u> VREs
- <u>Transparency</u>
- Change in the way we work change in the way we think
- <u>Transition</u> towards mega-science



LifeWatchGreece Project info

- Nov2012-Dec2015 (extension?)
- Almost 4 MEuros
- 49 Research Institutions and Academic Departments
- 400 participants from all over



What makes LifeWatchGreece?

- Human Network
- Physical Installations
- Equipment (mostly hardware)
- Software

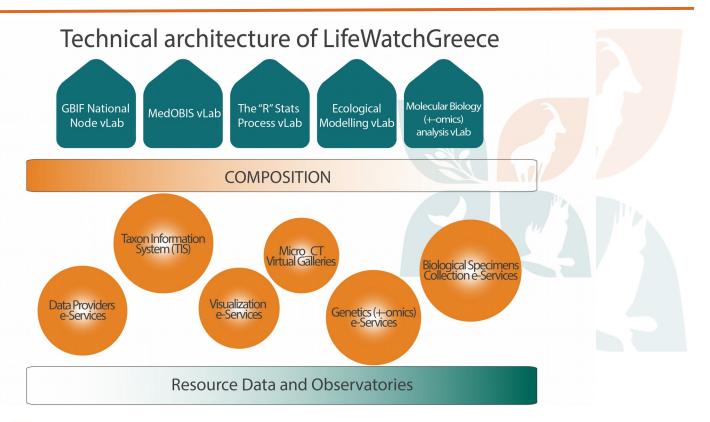


LifeWatchGreece Research (e-)Infrastructure



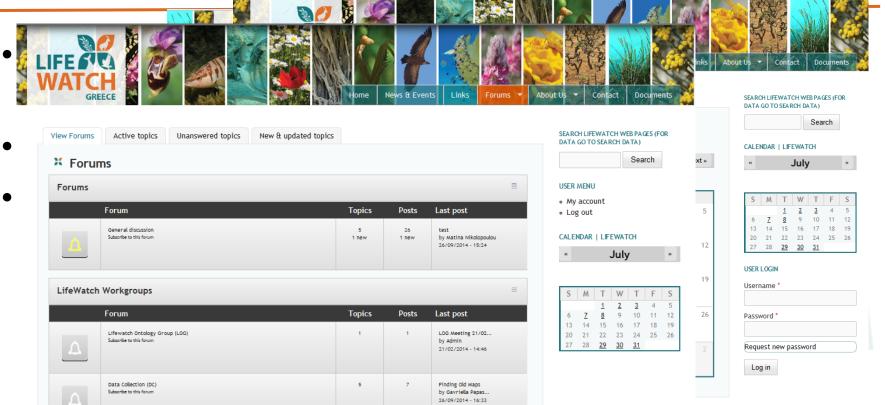


LifeWatchGreece Concept Architecture





Web portal





Experience / principles

- Not all researchers like the idea of data sharing: this is MY data
- Need to know the fair play side of the game: policy
- Incentives: data publicizing, data papers, data impact factor

- EU and MS are in favour of: <u>data is a capital produced by payers money</u>
- EU and MS are in favour of: data is a capital and can create jobs



Data Policy and Data Sharing Agreement

- http://www.lifewatchgreece.eu/?q=content/documents
- The landscape at the international and European scale
- LifeWatch ESFRI and LifeWatchGreece
- Why should I share my data?
- Why to support open access data?
- What kind of data can be submitted to LifeWatchGreece RI?
- Copyright, Creative Commons and related terms
- LifeWatchGreece General Terms of Use
- Rights and duties of the Data Providers & Data Owners
- Fair reuse of data published through LifeWatchGreece RI



Copyright, Creative Commons and related terms

- LifeWatchGreece RI licenses: LifeWatchGreece RI uses <u>Creative Commons</u> as a legal instrument to define the <u>usage rights of the</u> data. Creative commons is legally binding, simple to use, globally accepted and its licenses are both human-readable and machine-readable, the latter being especially important in the digital age.
- LifeWatchGreece RI data release: under two different conditions:
- 1) CC-Attribution (CC-BY, https://creativecommons.org/ licenses/by/3.0/): "You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use";
- 2) CC-Zero (waiver, https://creativecommons. org/publicdomain/zero/1.0/): "The person who associates a work with this deed has dedicated the work to the public domain by waiving all of his/her rights to the work worldwide under copyright law, including all related and neighboring rights, to the extent allowed by law. You can copy, modify,
- distribute and perform the work, even for commercial purposes, all without asking permission."
- Embargo: all data submitted to LifeWatchGreece RI can be subjected to an embargo period to be determined by the Data Provider and/or Data Owner.



Fair re-use of data published through LifeWatchGreece RI

The <u>Canadensys</u> norms (http://www.canadensys.net/) for data publication and use:

- Give credit where credit is due
- Be responsible
- Share knowledge
- Respect the data license



How can I participate?

- Explore the web site: http://www.lifewatchgreece.eu
- Inform us about your needs (methods of analysis, software, etc.)

Your messages to: info@lifewatchgreece.eu

Contribute and publish your own data and metadata



LifeWatchGreece - MedOBIS

Mediterranean node of Ocean Biogeographic Information System(OBIS)

- Virtual laboratory (vLab) in the LifeWatchGreece Research Infrastructure
- Platform offering a suite of applications to facilitate data gathering, storage & analysis devoted to Mediterranean Marine Biodiversity





MedOBIS - Infrastructure

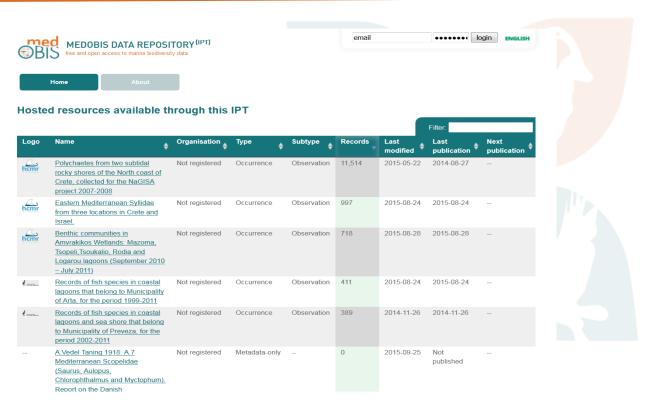
- Database (PostgreSQL database system) (50 datasets- some historical(Emodnet digitization)- 27 new are coming)
- IPT -GBIF Integrated Publishing Toolkit (http://ipt.medobis.eu/)
 (Darwin core schema –Follow Obis Schema- sampling events occurrences-measurements(extendedMeasurementorFacts) Expanding the Ocean Biogeographic
 Information System (OBIS) beyond species occurrences
- Geoserver -serves to MedOBIS viewer
- MedOBIS viewer (https://portal.lifewatchgreece.eu/) –medOBIS vLab



Watch LifeWatchGreece - MedOBIS IPT

The GBIF Integrated
Publishing Toolkit (IPT) is
a software
platform developed by the
Global
Biodiversity Information
Facility to facilitate efficient
biodiversity data sharing

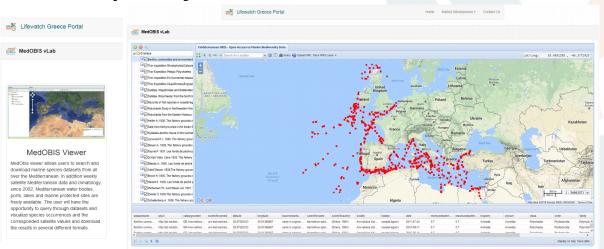
Data Harvesting provide to OBIS, and ultimately to GBIF





MedOBIS- Viewer

GeoExt & OpenLayers



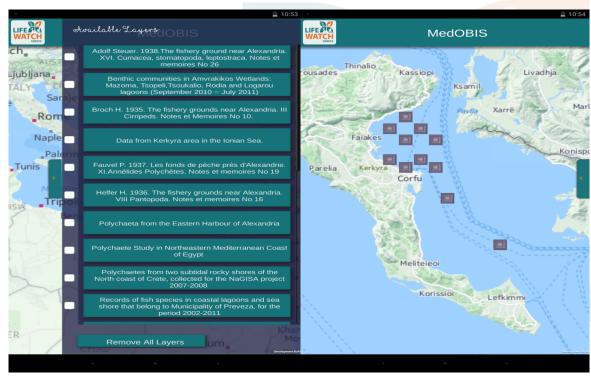
Users can visualize their own data in KML, publish and download the results in different formats (CSV, Geojson, WMS,WFS,KML).



LifeWatchGreece - More MedOBIS Apps

A MedOBIS mobile application is experimentally developed to run on smartphones and tablet computers.







- •Non destructive three-dimensional imaging technique similar to computer tomography used in hospitals, just on a much smaller scale
- •Samples of a few millimeters up to a size of a mouse, and structures in the range of a few microns (<0,5µm/pixel) can be seen in the images





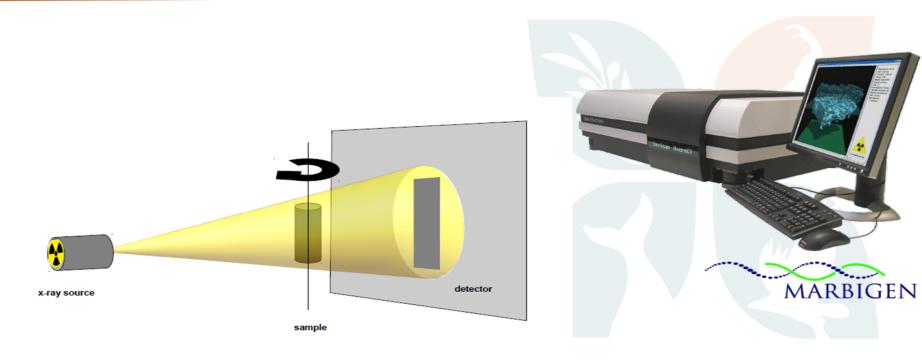




image acquisition





2. reconstruction



cross-sections

projection images

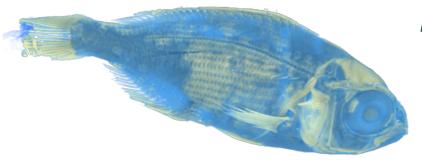


three-dimensional volume rendering



3. visualisation





Exterior parts: morphology







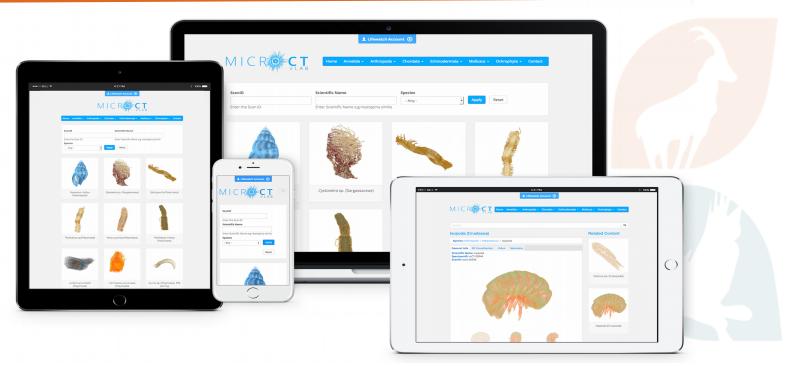


Micro-CT services are now available through http://microct.portal.lifewatchgreece.eu/





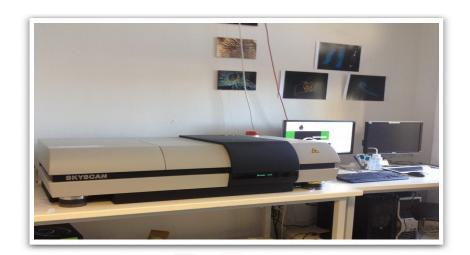






Micro-CT virtual galleries

- 650 scans have been created and for the initiation of this web service only 17 of them were uploaded
- The uploaded datasets belong to several marine organisms and they are fully annotated with metadata

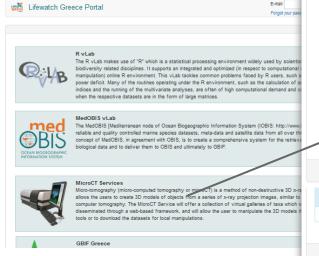


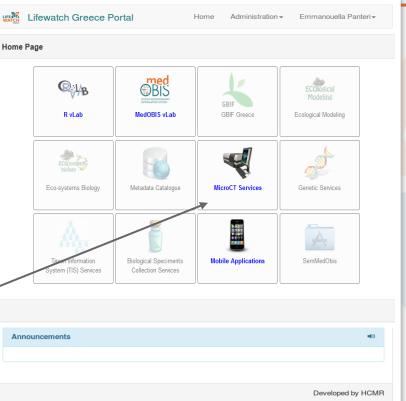
- 7 micro-CT datasets can be downloaded from the Dryad Digital Repository which is a repository system for several datatypes
- The remaining datasets can be shared through personal communication as the storage is still under construction



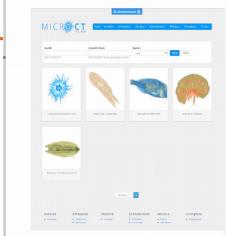
Accessed through LifewatchGreece Portal





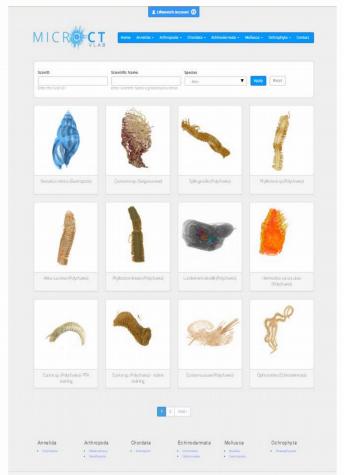




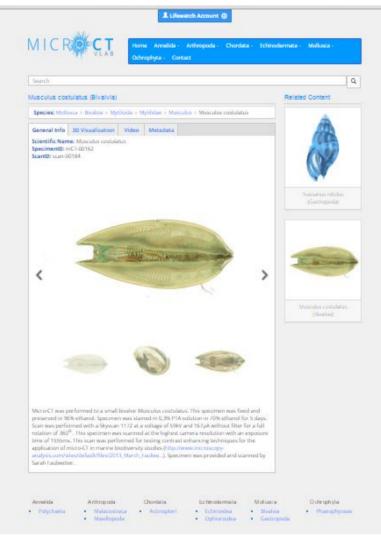


Scans are presented as a preview of images with the title of the dataset

Home Page





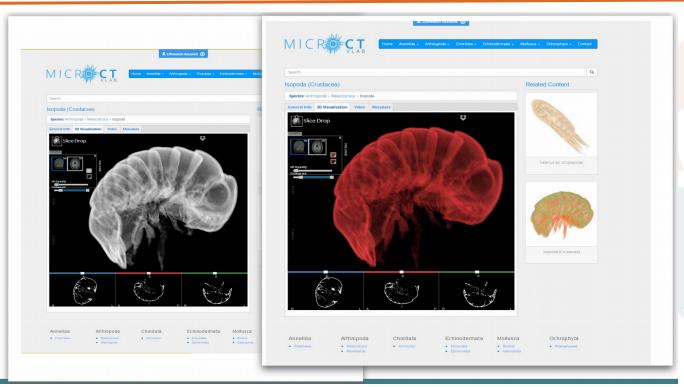


The micro-CT dataset

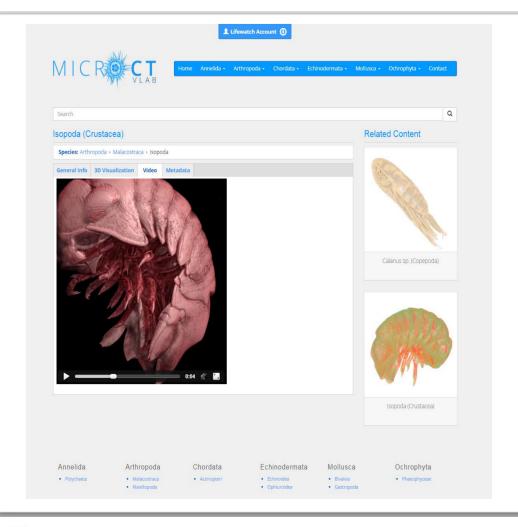
- Short description of the dataset
- Gallery of 3D images
- Related datasets



3D visualisation







video





1 Ufewatch Account (1)

Search	C
sopoda (Crustacea)	Related Content
Species: Arthropoda » Malacostraca » Isopoda	
General Info 3D Visualization Video Metadata	B.
MicroCT Scanning Results Found	93
Result 1 (Related Dataset):	1000
Specimen ID: http://lifewatch.gr/entity/specimen/mct-00001	
Specimen Name: mCT-00001	
Scanning ID: http://lifewatch.gr/entity/digitizationProcess/scan-00001	The second second
Product ID: http://lifewatch.gr/entity/dataObject/scan-00001.zip	
Product Name: scan-00001.zip	Calanus sp. (Copepoda)
Device ID: http://lifewatch.gr/entity/device/skyscan_1172	
Device Name: SkyScan 1172	
Actor ID: http://lifewatch.gr/entity/actor/sarah_faulwetter	
Actor Name: Sarah Faulwetter	
Date: 2013-02-18	
Dataset Name: microCT_ElectronicLog	
Dataset ID: http://lifewatch.gr/entity/dataset/microct_electroniclog	6811119
Result 2 (Related Dataset):	
Specimen ID: http://lifewatch.gr/entity/specimen/paraehlersia_ferrugina-cala-20c_07	
Specimen Name: Paraehlersia_ferrugina-CALA-20C_07	,
Scanning ID: http://lifewatch.gr/entity/digitizationProcess/scan-00011	
Product ID: http://lifewatch.gr/entity/dataObject/scan-00011.zip	Isopoda (Crustacea)
Product Name: scan-00011.zip	
Device ID: http://lifewatch.gr/entity/device/skyscan_1172	
Device Name: SkyScan 1172	
Actor ID: http://lifewatch.gr/entity/actor/sarah_faulwetter	
Actor Name: Sarah Faulwetter	
Date: 2012-04-25	
Dataset Name: microCT_ElectronicLog	

Metadata





Big challenges

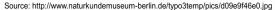
- Huge storage capacity demand
- Downloading bandwidth
- 3D Pattern recognition solutions



Big changes

the idea







type material



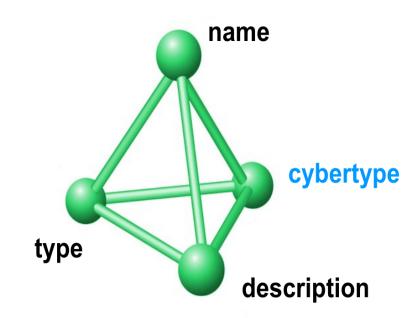
cybertype

Godfray, HCJ (2007) Linnaeus in the information age. *Nature* 446: 259–260



Big changes: from phenetics to phenomics

... to the "taxonomic tetrahedron"?







The first idea!

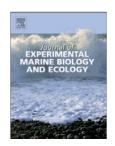
Journal of Experimental Marine Biology and Ecology 366 (2008) 184-186



Contents lists available at ScienceDirect

Journal of Experimental Marine Biology and Ecology

journal homepage: www.elsevier.com/locate/jembe



All animals are equal, but some animals are more equal than others

R.M. Warwick *, P.J. Somerfield

Plymouth Marine Laboratory, Prospect Place, West Hoe, Plymouth, PL1 3DH, UK



What was the concept?

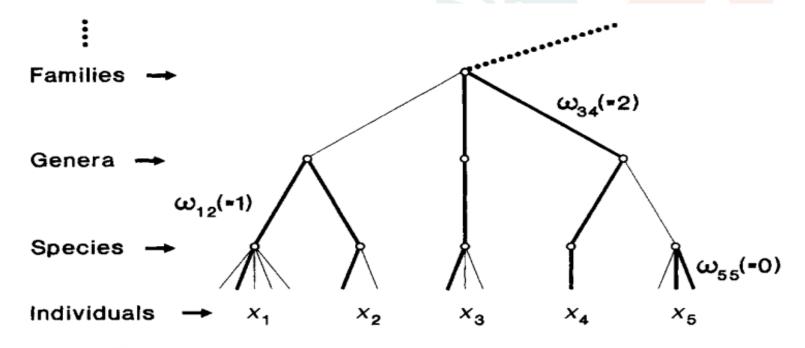
1. "If the number of animal species is to be used as a measure of 'biodiversity', or if distributions of species among taxa of higher rank are to be used to infer evolutionary or ecological patterns, then we need to know whether animal phyla are consistently subdivided in such a way that each species represents an equal division of life's diversity.

2. It is widely assumed, intuitively, that the traditional Linnean classification of marine animals is inconsistent between different major groups. We demonstrate formally that this is the case."



What was the concept?

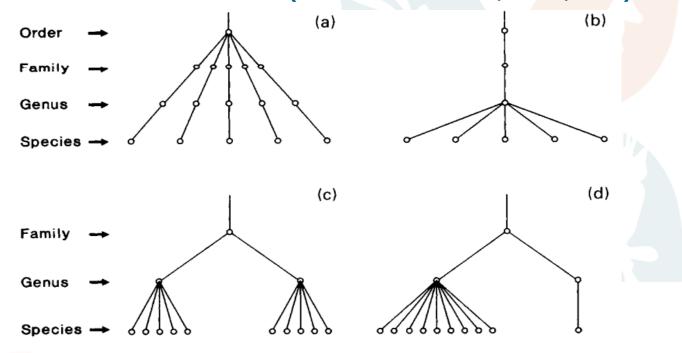
The taxonomic distinctness indices (Clarke & Warwick, 2998; JAE):





What was the concept?

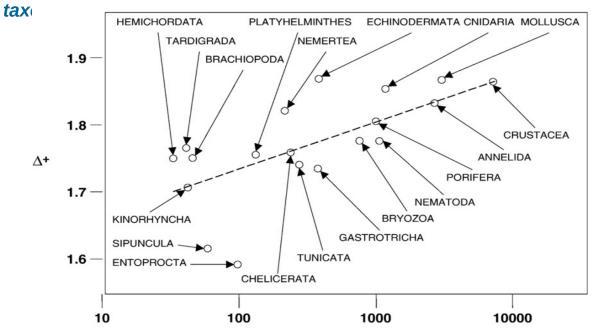
The taxonomic distinctness indices (Clarke & Warwick, 2998; JAE):





What was the result?

3. "The value of average taxonomic distinctness Δ^+ is shown to vary considerably between phyla. There is a highly significant relationship between the number of species within a phylum and the average distance through the





What were the conclusions?

- 4. "There is a highly significant relationship between the number of species within a phylum and the average distance through the taxonomic hierarchy between those species.
- **5.** This implies that larger phyla are broken up into relatively small units at higher taxonomic levels. Interestingly, this occurs independently of the perceived taxonomic difficulty within phyla.
- **6. Species number** is therefore a **poor unit of currency** for **evaluating biodiversity**, and studies which **infer patterns** using **distributions of, or ratios between, higher taxa** through time should take **phyletic differences** into account."



What is the current concept?

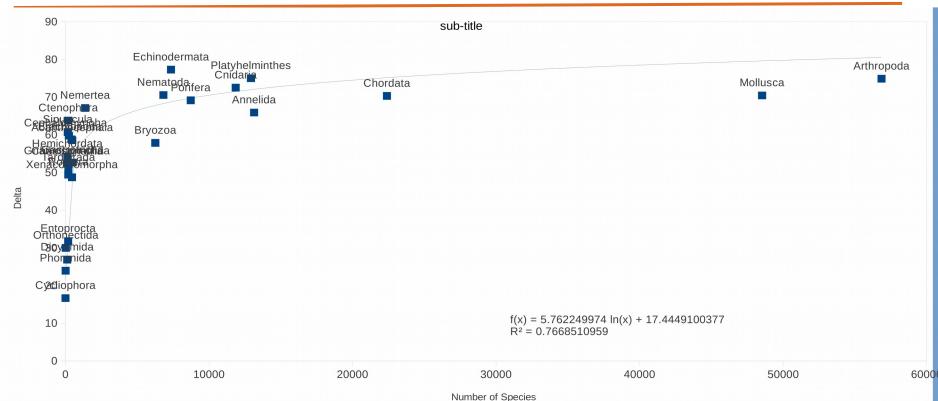
1. Pretty much the same but now we scale up:

2. From the UK inventory to WoRMS

3. Include both average and variation in taxonomic distinctness (Δ^+ , Λ^+) to infer from both the average distance **length** but also from the (un)equal distribution of species to higher categories (**width**) of the WoRMS phylogenetic/ taxonomic tree.



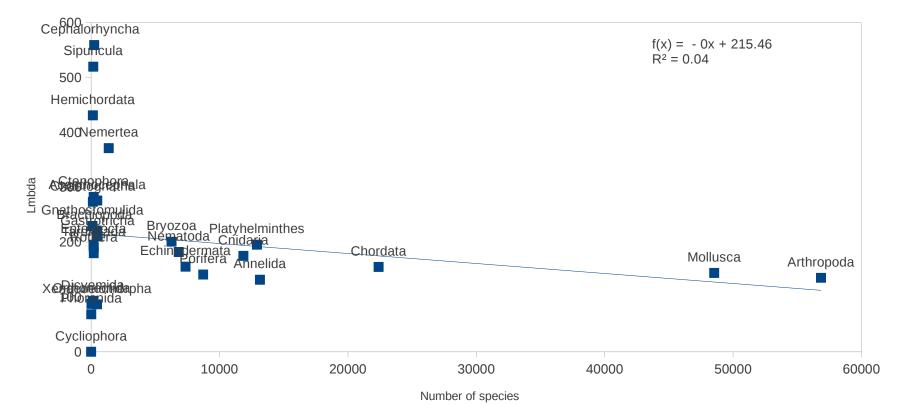
What are the results?





What are the results?

sub-title





What are the results?

- 1. The value of average taxonomic distinctness Δ^+ still varies considerably between phyla but much less so. Again, there is a highly significant relationship between the number of species within a phylum and the average distance through the taxonomic hierarchy between those species but this relationship becomes logarithmic now.
- 2. The value of variation in taxonomic distinctness Λ⁺ varies enormously between phyla, especially between those with less than 5,000 species. There is still a positive but weak linear relationship between the number of species within a phylum and the average spread of the species into higher categories.



What might these results atch mean?

- **1.** Again, "larger phyla are broken up into relatively small units at higher taxonomic levels. Interestingly, this occurs independently of the perceived taxonomic difficulty within phyla".
- 2. The two biodiversity indices bring complementary patterns when applied to the entire WoRMS tree.
- **3.** Phyla also differ in the horizontal dimension of their architecture, which now needs to be investigated in the lower taxonomic categories: e.g. classes, orders and families, within each phylum.
- **4.** We currently **need** not only to know whether species number is a **poor unit of currency** for **evaluating biodiversity,** but also if higher categories can serve as such currency.
- 5. Do these trends occur in other phylogenetic systems (CBOL, ATOL)? If not what are the differences and why?



Next steps are....

1. Apply the same methodology to the lower taxonomic categories.

2. Test the **phylogenetic trees** produced by **CBOL** and **ATOL** for the same hypothesis and make comparisons with the present findings (this will require an adjustment of the WoRMS tree).

LifeWatclWhat a collaborative VRE might look like?



















CMBR **MOUNT KRIPIS**



GREECE

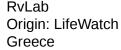


Sorting and Integration layer











Micro-CT vLab Origin: LifeWatch Origin: IMBBC Greece



Genetic Analysis vLab Taxonomic Greece



Backbone vLab Origin: LifeWatch Belgium



QC services Origin: LifeWatch Belgium



Data validation and MarineRegions gazetteer services Origin: LifeWatch Belgium

Vlabs and eServices layer



Thank you for your attention Special tanks to W. Los for the LifeWatch slides

http://www.lifewatchgreece.eu nfo@lifewatchgreece.eu