Cyberspecimens

from creation
to curation
and publication

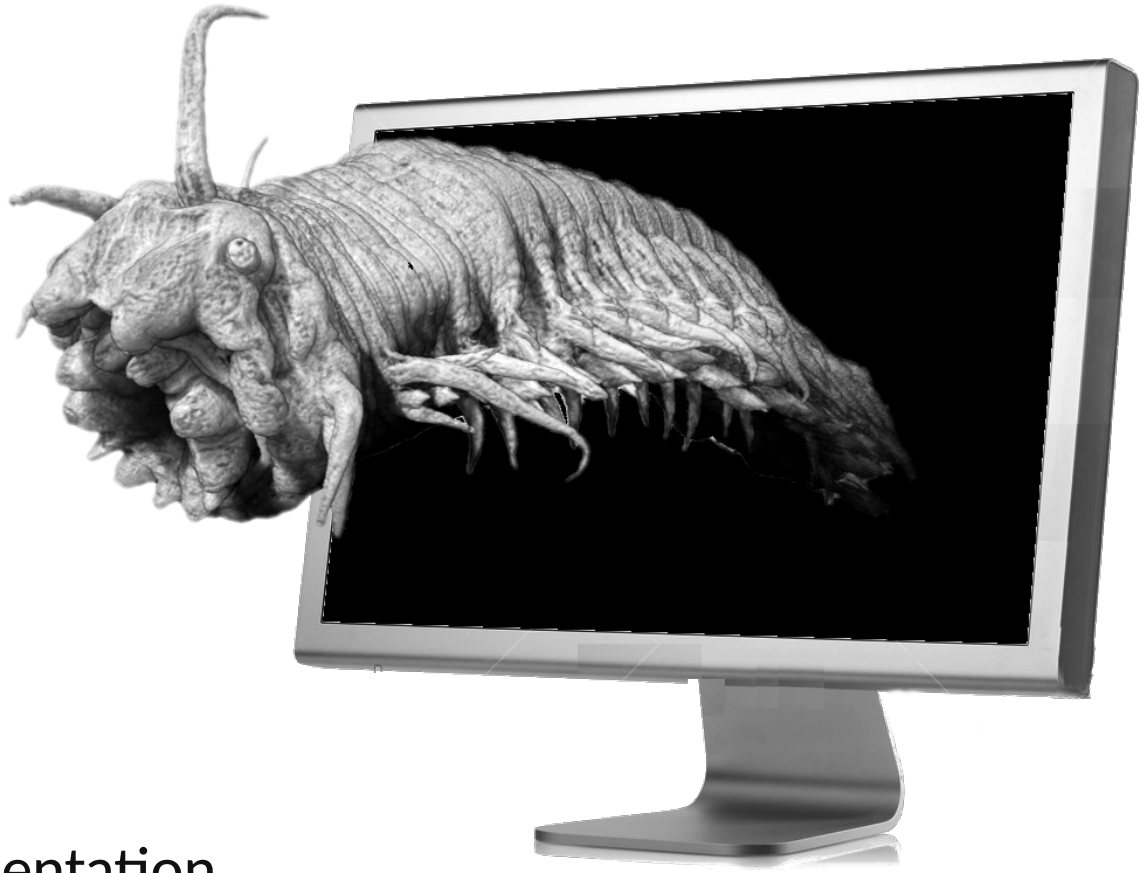
Sarah Faulwetter

@sarsifa

Cyberspecimen

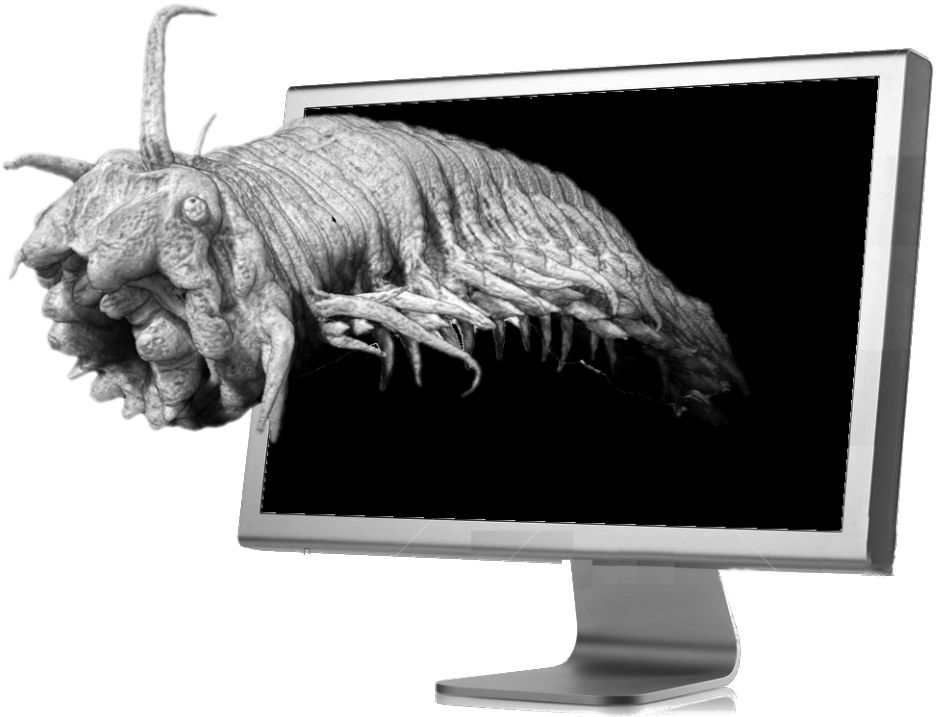
Cybertype

Avatar



Accurate, virtual (3D) representation
of the physical specimen

By Hannes Grobe/AWI CCBY 3.0 via Wikimedia Commons



type material → cybertype

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

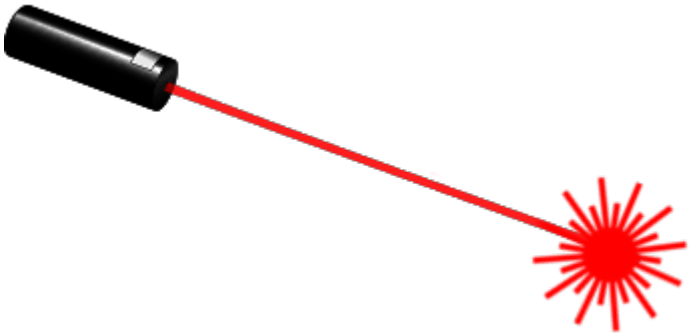
Volume data (internal & external structures)

X-ray tomography / micro-CT

Magnet Resonance Imaging (MRI)

Surface data (external structures)

Photogrammetry

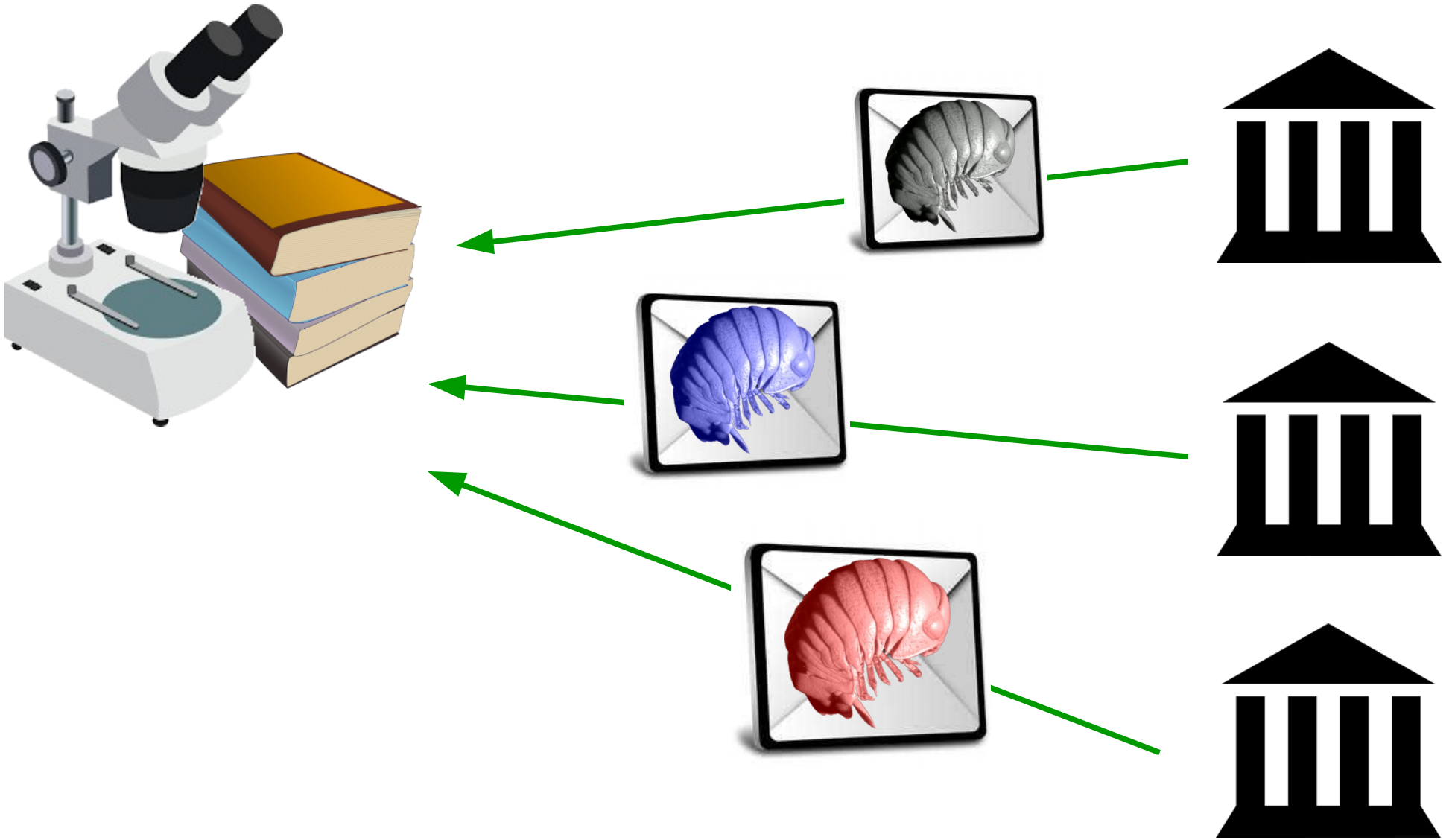


Laser surface scanning

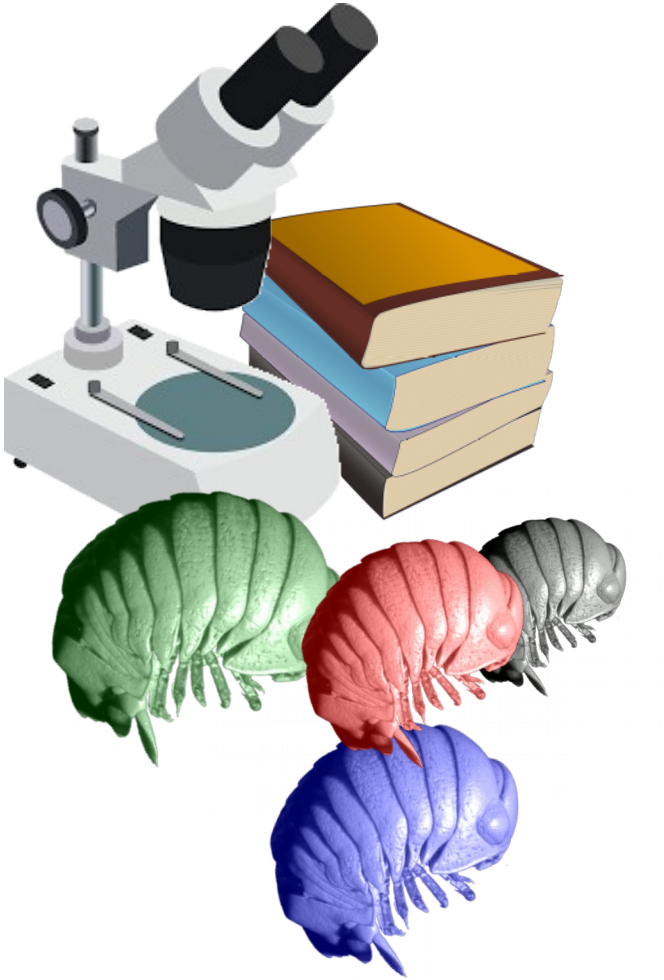
Why cyberspecimens?

The taxonomic bottleneck

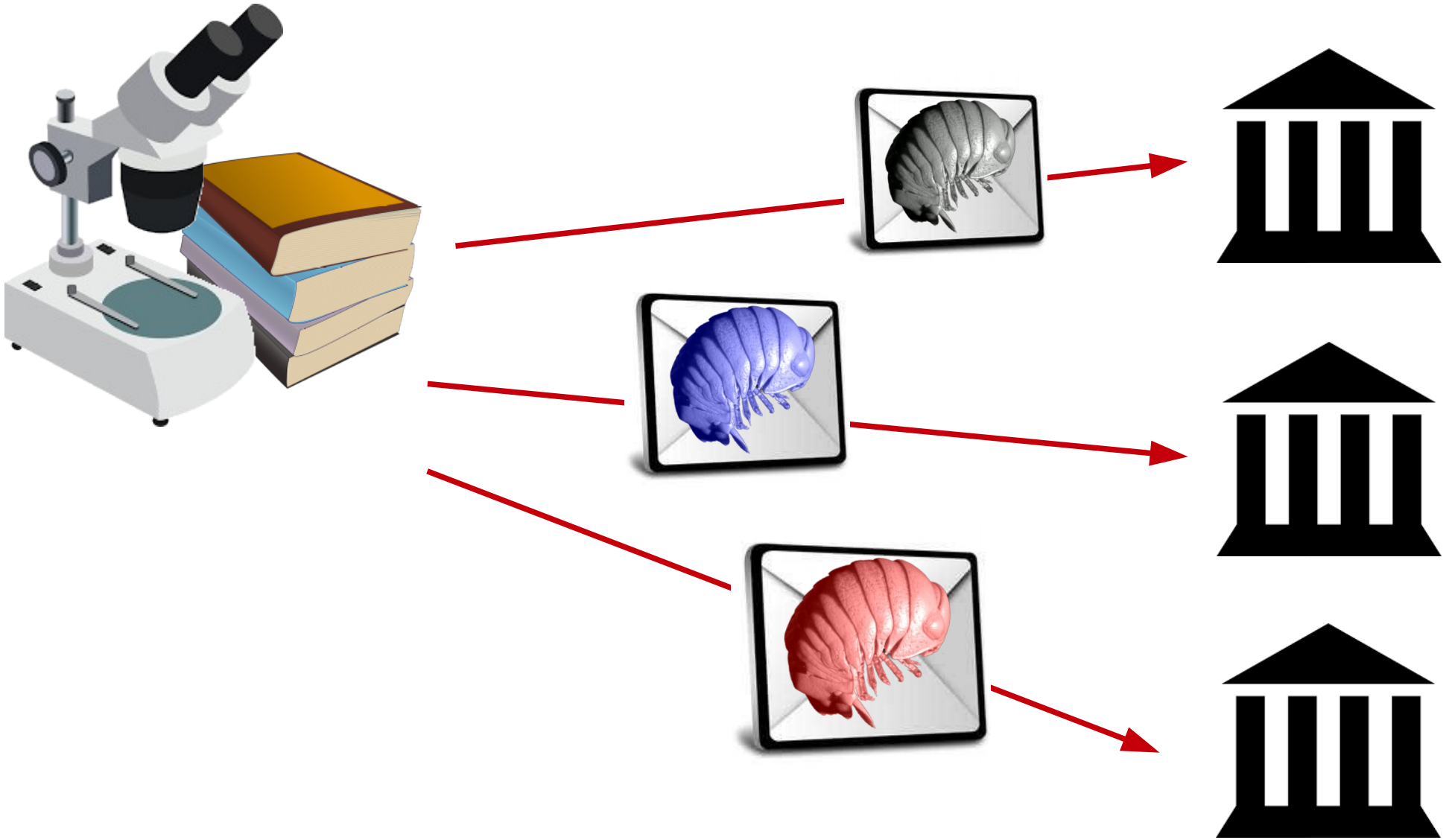
The taxonomic bottleneck



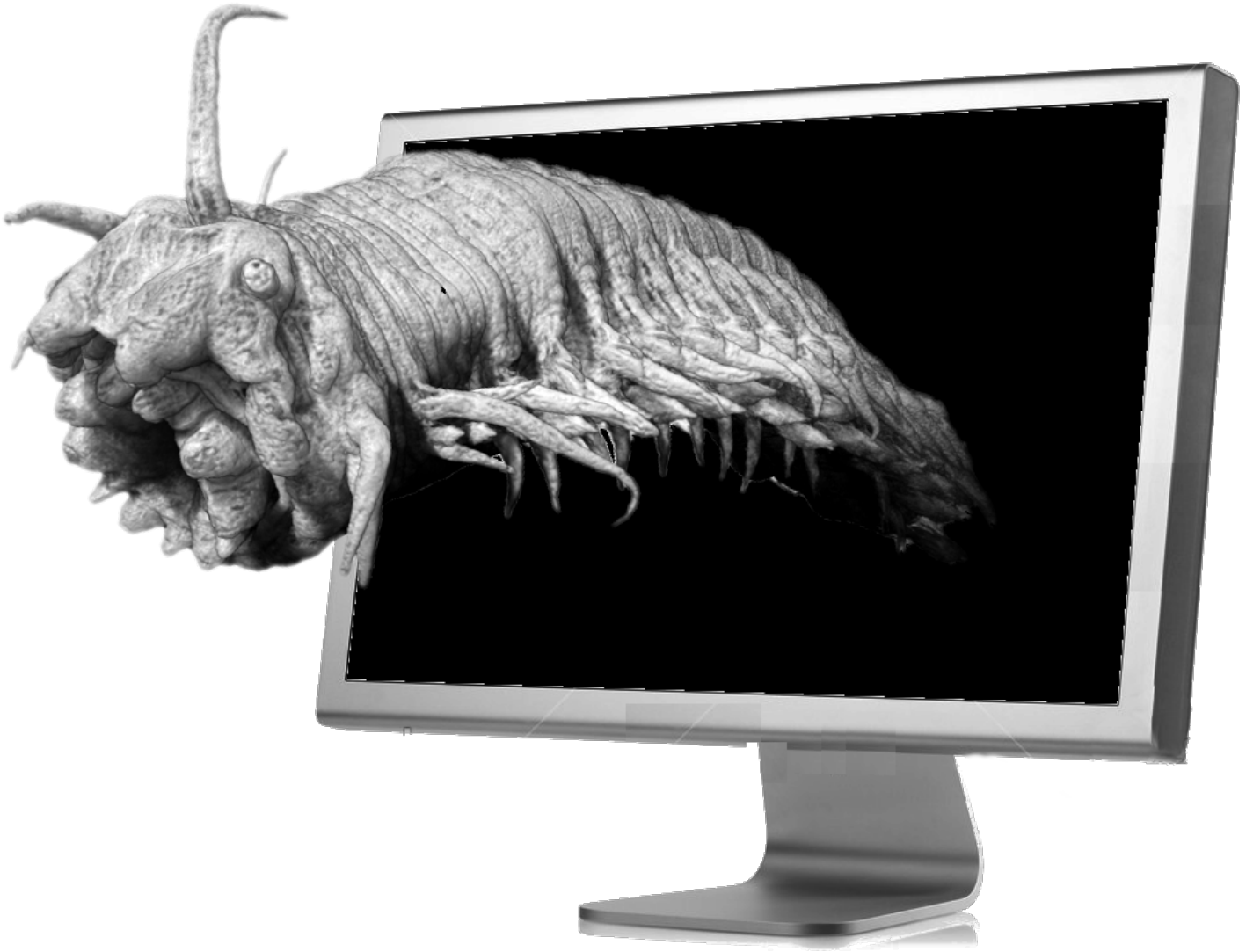
The taxonomic bottleneck



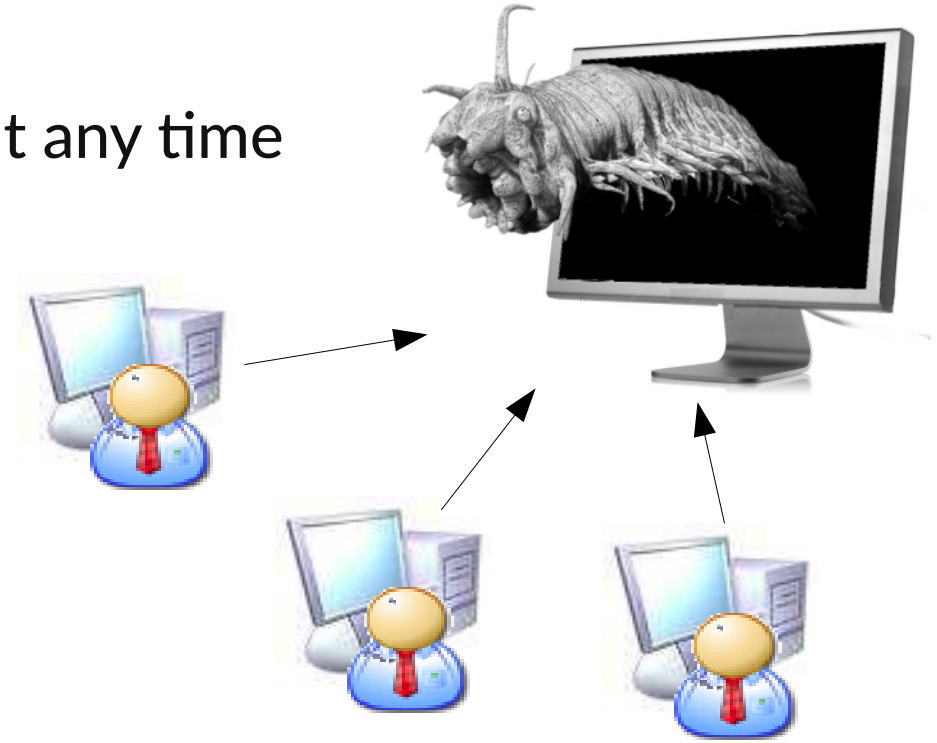
The taxonomic bottleneck



Cybertype

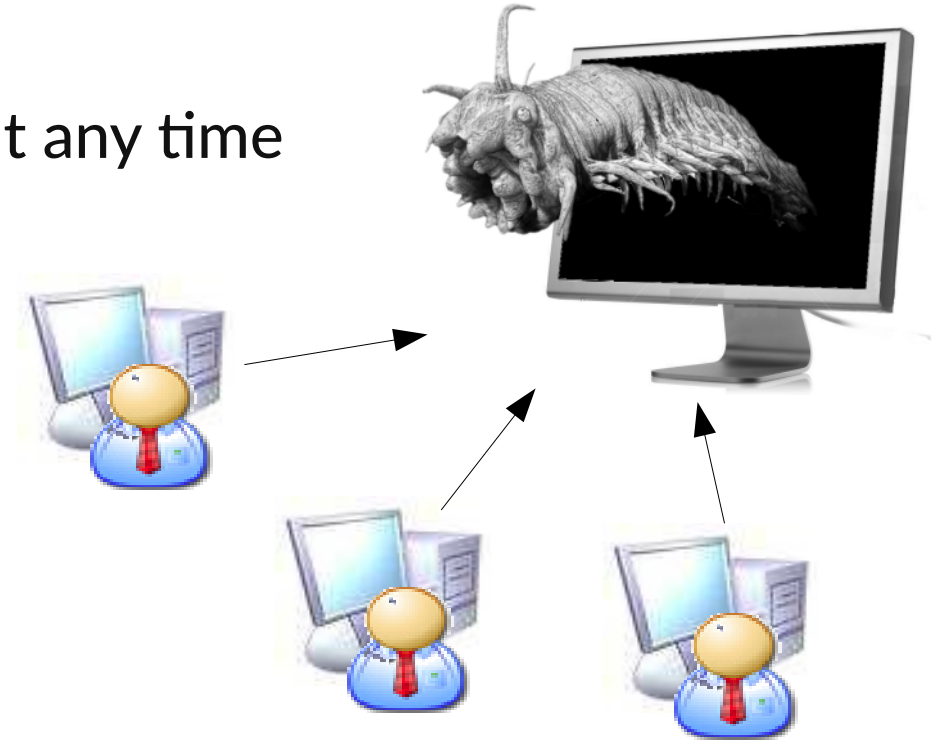


instant access to museum material at any time



...simultaneously, by multiple researchers from any place

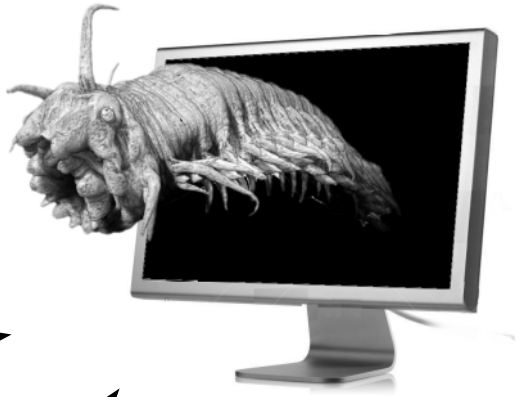
instant access to museum material at any time

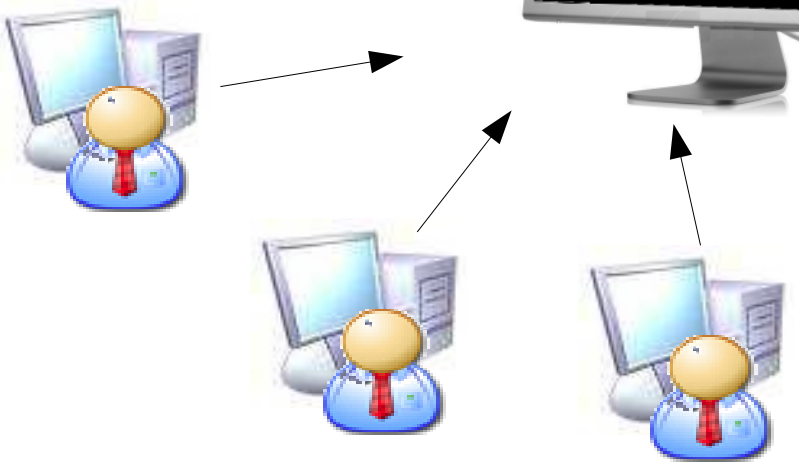


...simultaneously, by multiple researchers from any place

protection of valuable specimens from damage or loss

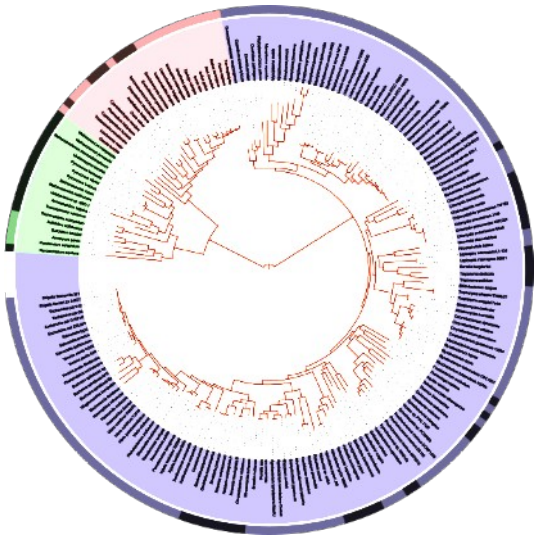
instant access to museum material at any time





...simultaneously, by multiple researchers from any place

protection of valuable specimens from damage or loss



digital format → usage by computers
→ new types of analyses

A cyberspecimen should...

I provide information of the same accuracy and reliability as the physical specimen

II be created without affecting the morphological, anatomical and molecular identity of the physical specimen

III be linked to the original specimen (which can be consulted if in doubt)

IV be retrievable

A cyberspecimen should...

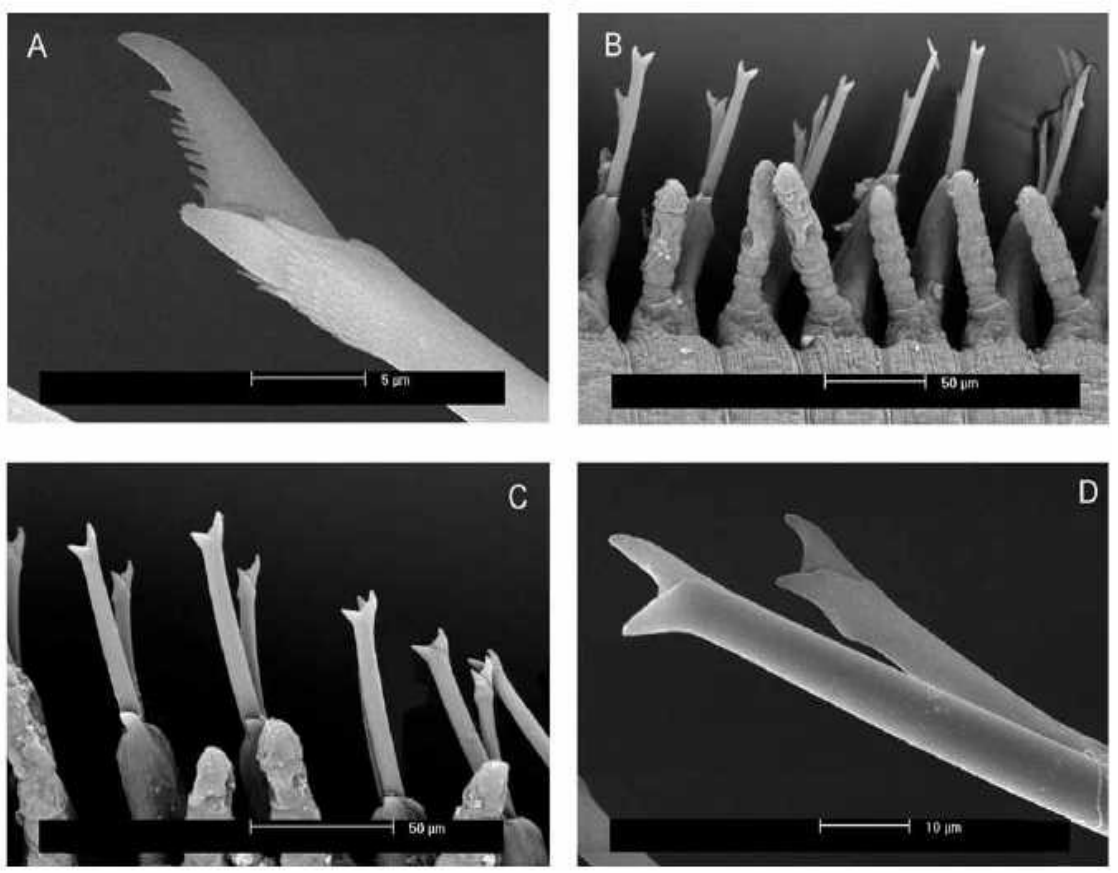
provide information of the same accuracy and reliability as the physical specimen

II be created without affecting the morphological, anatomical and molecular identity of the physical specimen

III be linked to the original specimen (which can be consulted if in doubt)

IV be retrievable

Scanning Electron Microscope (SEM)

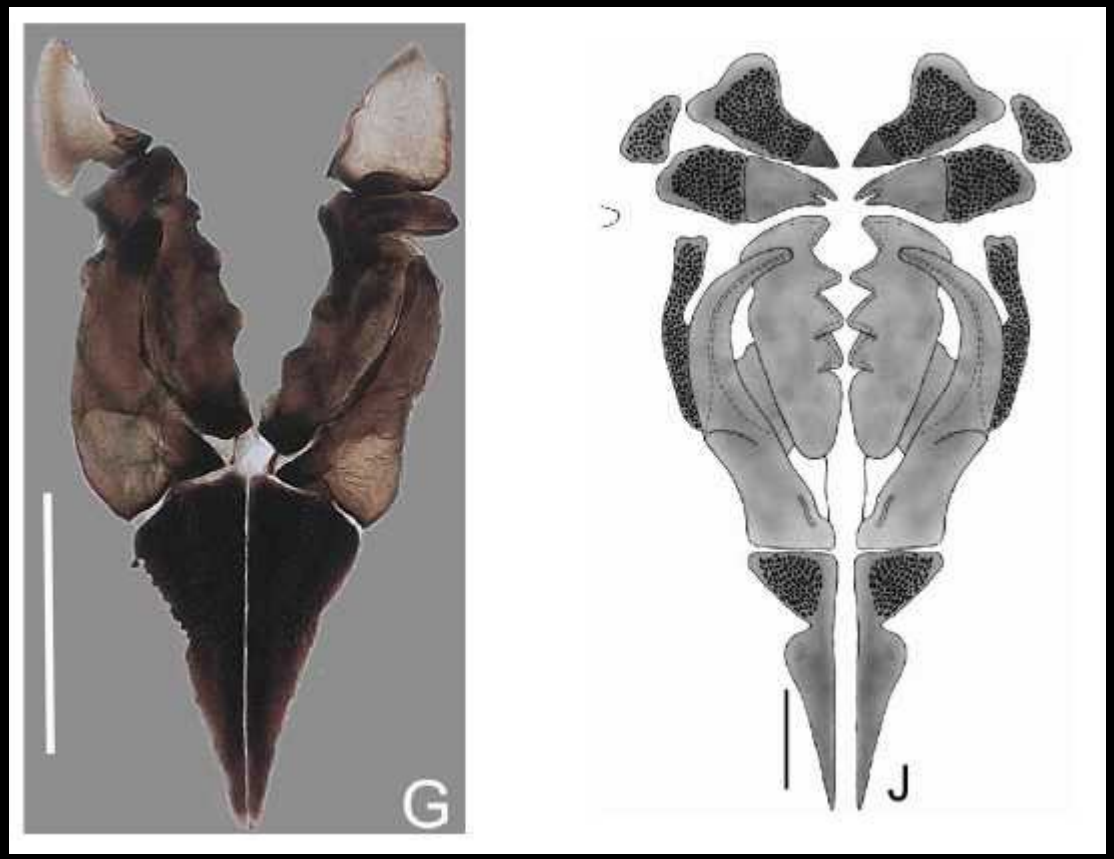


Source: Aguado et al. 2008, Zootaxa

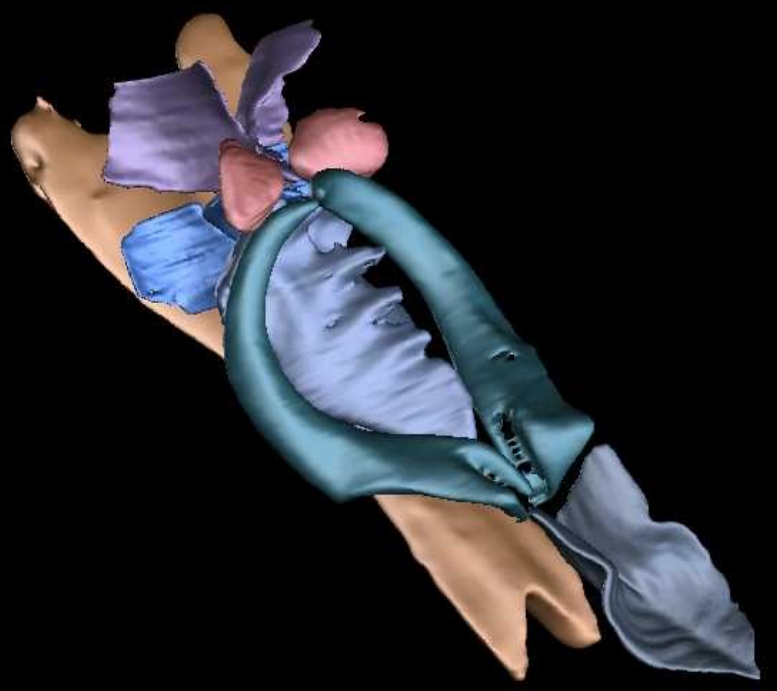
Micro-CT

Syllis gracilis (Annelida)

Light microscope and drawing



Micro-CT (surface model)



Lumbrineris latreilli (Annelida)

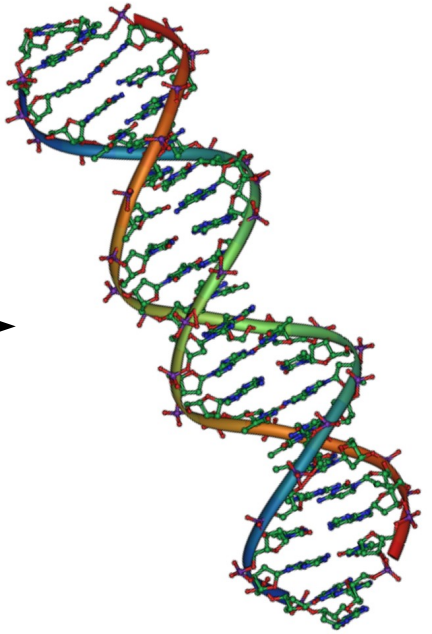
A cyberspecimen should...

provide information of the same accuracy and reliability as the physical specimen

be created without affecting the morphological, anatomical and molecular identity of the physical specimen

III be linked to the original specimen (which can be consulted if in doubt)

IV be retrievable




7% Lugol's solution
1 fl. oz.
Purified water, potassium


SPI-CHEM
SPI-CHEM
SPI-CHEM
Phosphotungstic
Acid
LOT# 1201102
Batch# X06A018


ALDRICH
Chemistry
440191-100ML
Lot # SHBB7795V
Hexamethyldisilazane
reagent grade, ≥99%

A cyberspecimen should...

provide information of the same accuracy and reliability as the physical specimen

be created without affecting the morphological, anatomical and molecular identity of the physical specimen

be linked to the original specimen (which can be consulted if in doubt)

IV

be retrievable

A cyberspecimen should...

provide information of the same accuracy and reliability as the physical specimen

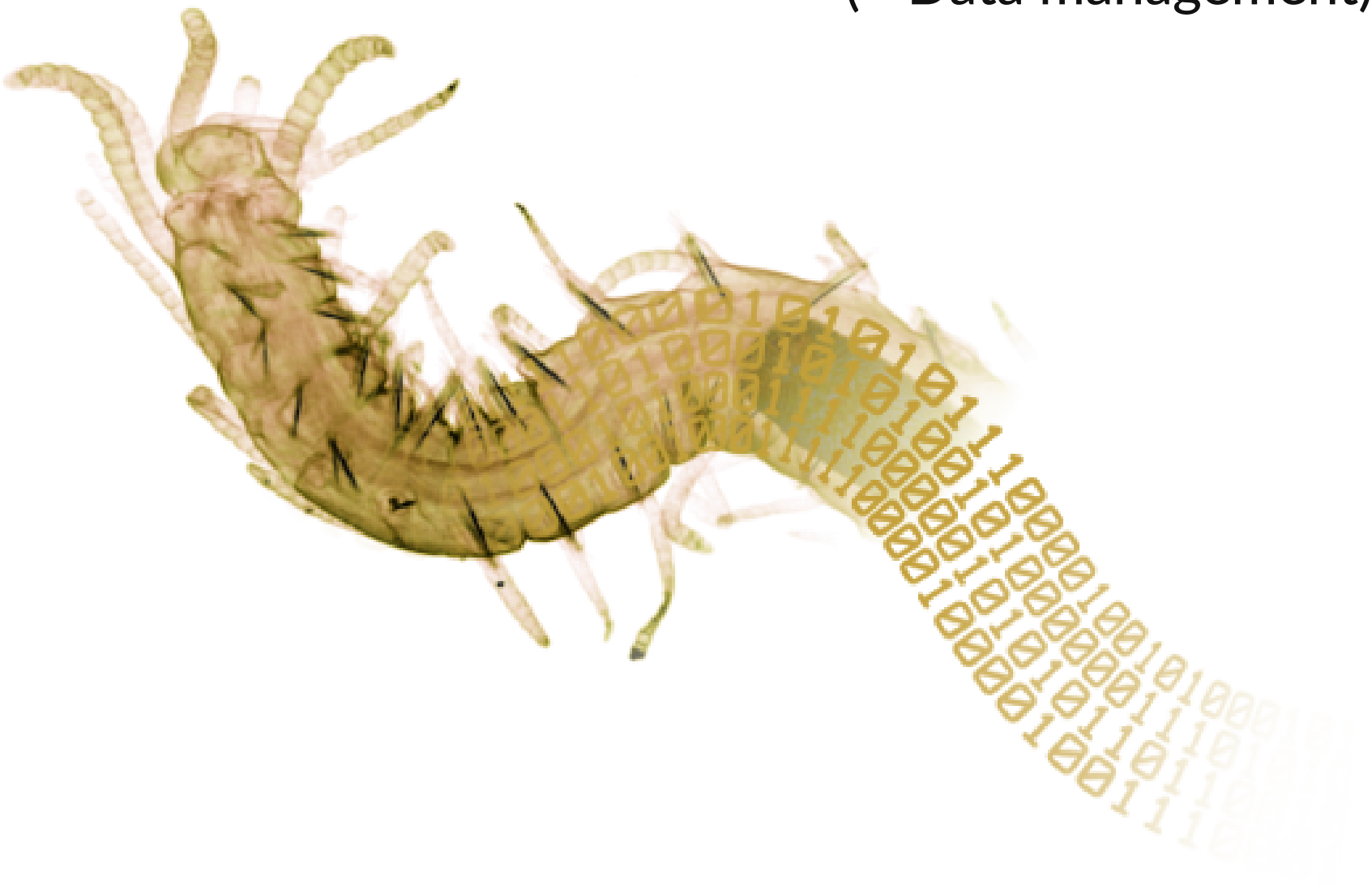
be created without affecting the morphological, anatomical and molecular identity of the physical specimen

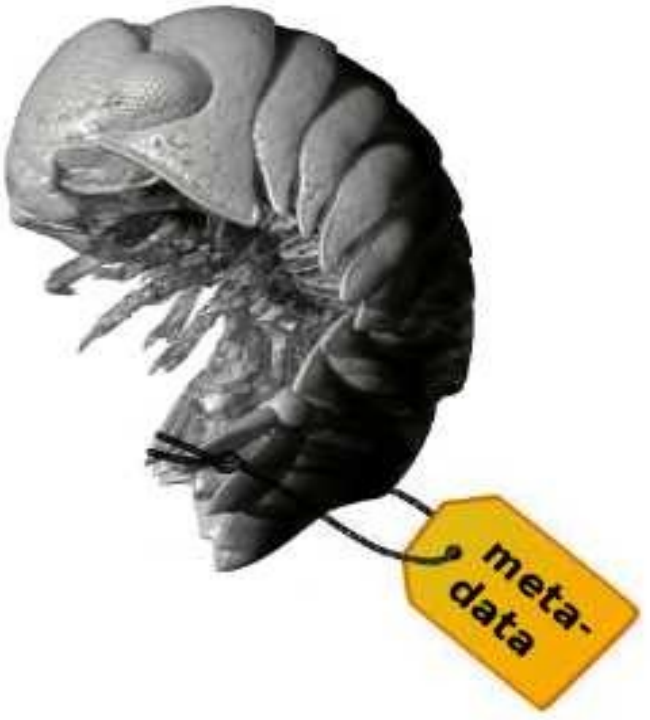
be linked to the original specimen (which can be consulted if in doubt)

be retrievable

Curation

(= Data management)





metadata / annotation system

Purpose of metadata

retrievability and discoverability

retrievability and discoverability

promote re-use of data

-> avoid duplication of efforts

promote re-use of data

-> avoid duplication of efforts

retrievability and discoverability

data interpretation

retrievability and discoverability

promote re-use of data

-> avoid duplication of efforts

data interpretation

process of information creation

can be understood and repeated

assess fitness for use

retrievability and discoverability

promote re-use of data
-> avoid duplication of efforts

data interpretation

process of information creation
can be understood and repeated

assess fitness for use

location, accessibility and terms of use

retrievability and discoverability

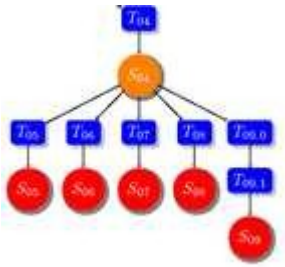
promote re-use of data
-> avoid duplication of efforts

data interpretation

process of information creation
can be understood and repeated

assess fitness for use

location, accessibility and terms of use



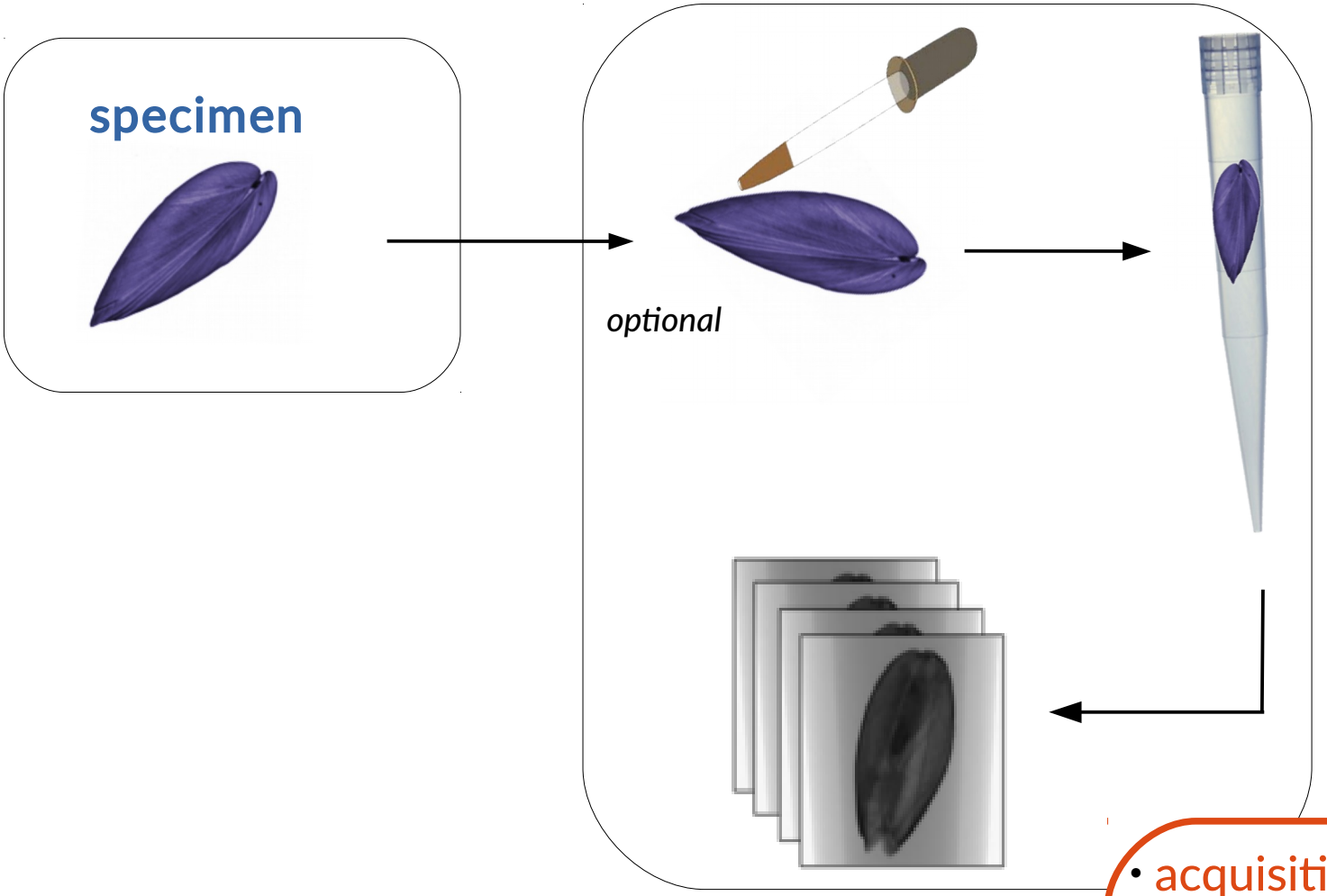
relationship with other resources

What and how should I record?

No agreed standards or guidelines for natural history 3D data

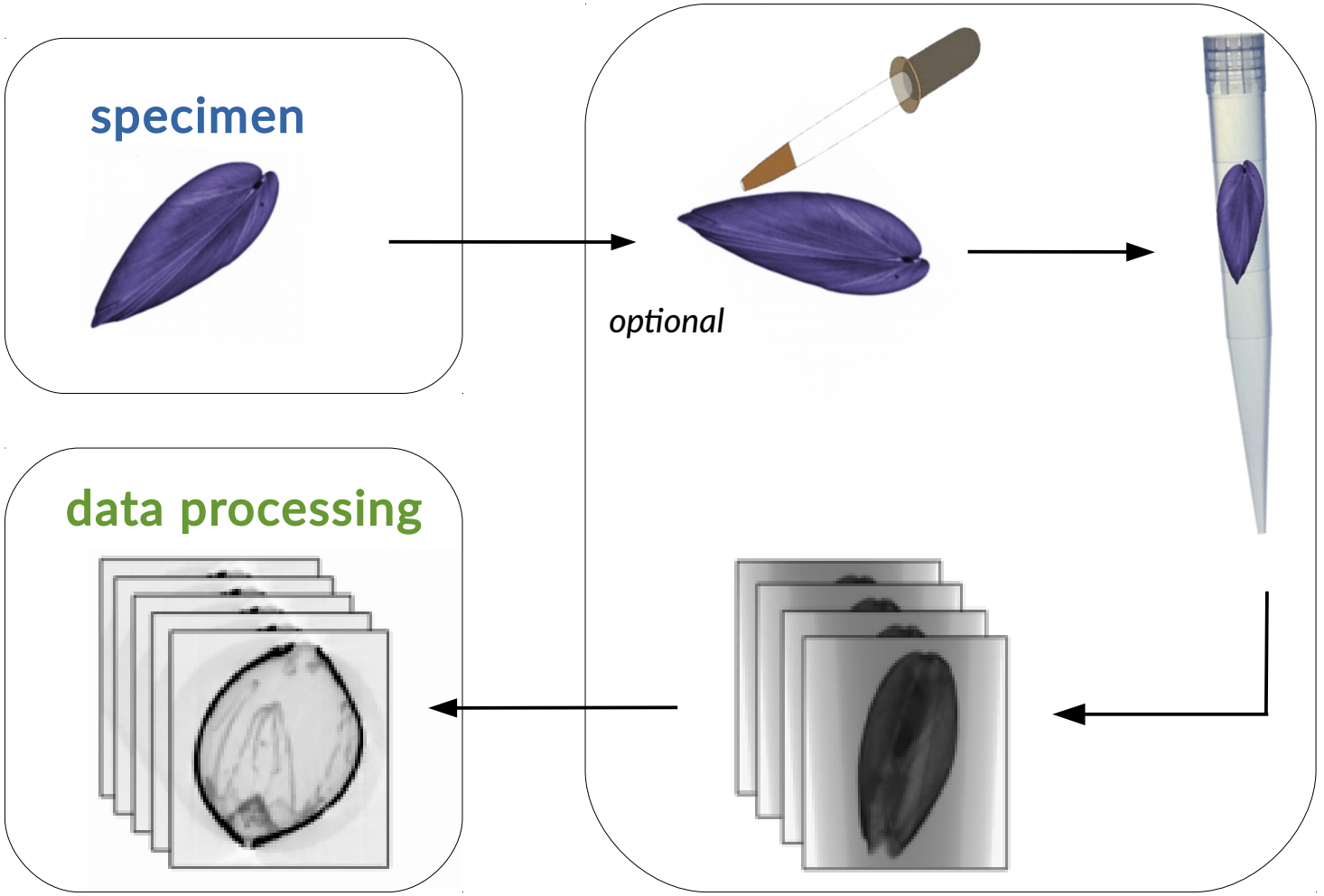
specimen

- Specimen ID
- Specimen provider
- Taxon name
- Specimen description
- Size (mm)
- Fixation medium
- Preservation medium

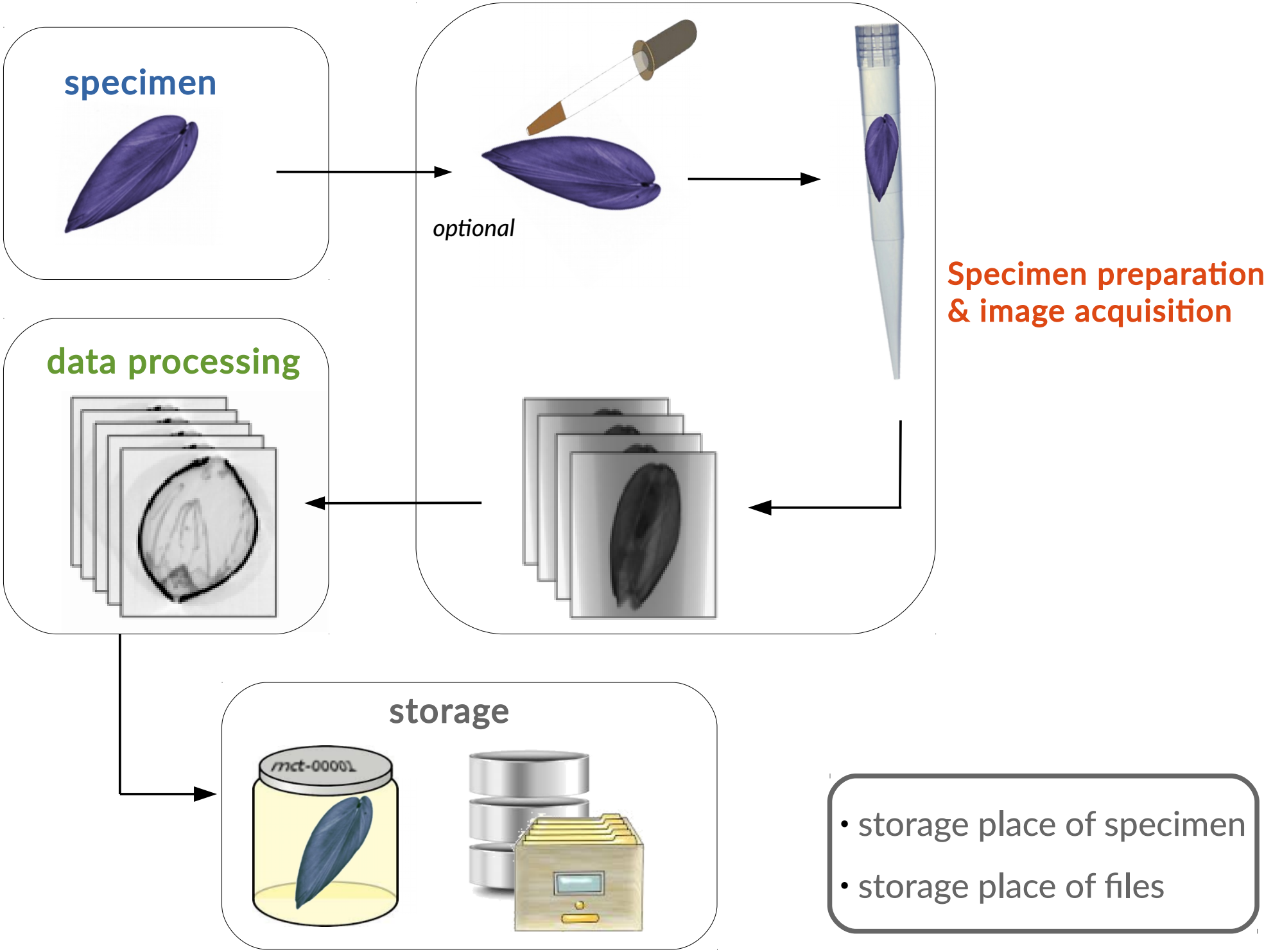


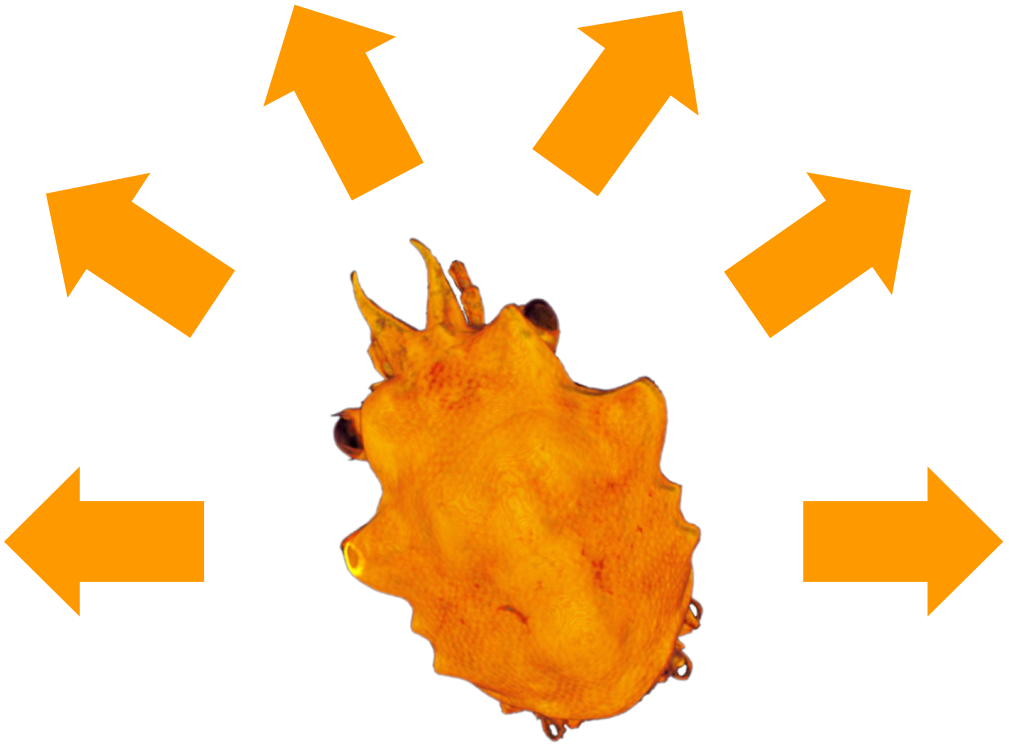
Specimen preparation & image acquisition

- acquisition ID
- [pre-treatment details]
- [sample holder & medium]
- person performing acquisition
- acquisition date
- scope of acquisition
- acquisition parameters



- scope of processing
- person processing data
- processing date
- processing steps / details


specimen
optional
Specimen preparation
& image acquisition
data processing
storage
mct-00001
• storage place of specimen
• storage place of files

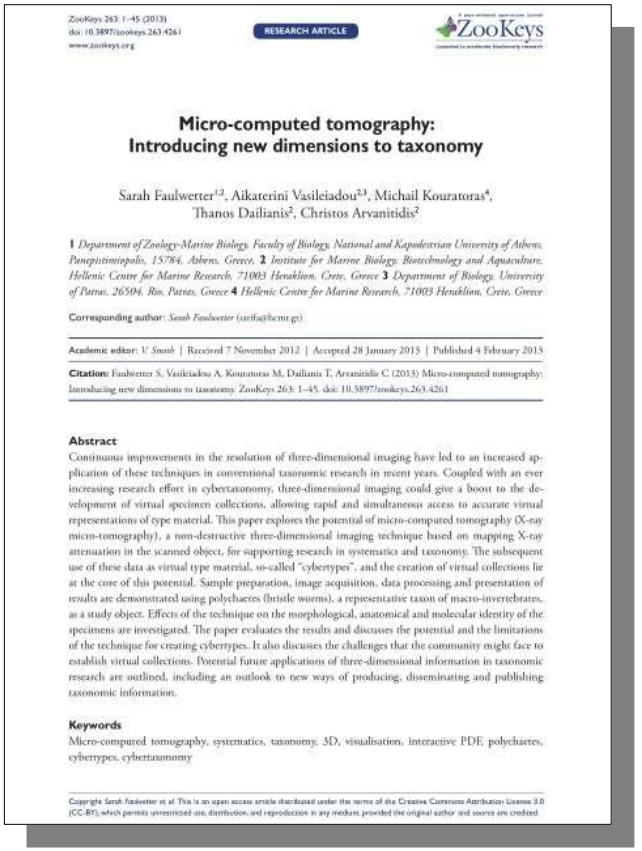


publication / dissemination

Purpose of data publication

(archival, deposition)

part of the
institute's mission



ZooKeys 263: 1–45 (2013)
doi: 10.3897/zookeys.263.4261
www.zookeys.org

RESEARCH ARTICLE

ZooKeys

Micro-computed tomography: Introducing new dimensions to taxonomy

Sarah Faulwetter[1,2], Aikaterini Vasileiadou[2,3], Michail Kouratoras[4], Thanos Dailianis[2], Christos Arvanitidis[2]

1 Department of Zoology-Marine Biology, Faculty of Biology, National and Kapodestrian University of Athens, Panepistimiopolis, 15784, Athens, Greece. 2 Institute for Marine Biology, Biotechnology and Aquaculture, Hellenic Centre for Marine Research, 71003 Heraklion, Crete, Greece 3 Department of Biology, University of Patras, 26504, Rio, Patras, Greece 4 Hellenic Centre for Marine Research, 71003 Heraklion, Crete, Greece

Corresponding author: *Sarah Faulwetter* (sarifa@hcmr.gr)

Academic editor: *V. Smith* | Received 7 November 2012 | Accepted 28 January 2013 | Published 4 February 2013

Citation: Faulwetter S, Vasileiadou A, Kouratoras M, Dailianis T, Arvanitidis C (2013) Micro-computed tomography: Introducing new dimensions to taxonomy. ZooKeys 263: 1–45. doi: 10.3897/zookeys.263.4261

Abstract

Continuous improvements in the resolution of three-dimensional imaging have led to an increased application of these techniques in conventional taxonomic research in recent years. Coupled with an ever increasing research effort in cybertaxonomy, three-dimensional imaging could give a boost to the development of virtual specimen collections, allowing rapid and simultaneous access to accurate virtual representations of type material. This paper explores the potential of micro-computed tomography (X-ray micro-tomography), a non-destructive three-dimensional imaging technique based on mapping X-ray attenuation in the scanned object, for supporting research in systematics and taxonomy. The subsequent use of these data as virtual type material, so-called "cybertypes", and the creation of virtual collections lie at the core of this potential. Sample preparation, image acquisition, data processing and presentation of results are demonstrated using polychaetes (bristle worms), a representative taxon of macro-invertebrates, as a study object. Effects of the technique on the morphological, anatomical and molecular identity of the specimens are investigated. The paper evaluates the results and discusses the potential and the limitations of the technique for creating cybertypes. It also discusses the challenges that the community might face to establish virtual collections. Potential future applications of three-dimensional information in taxonomic research are outlined, including an outlook to new ways of producing, disseminating and publishing taxonomic information.

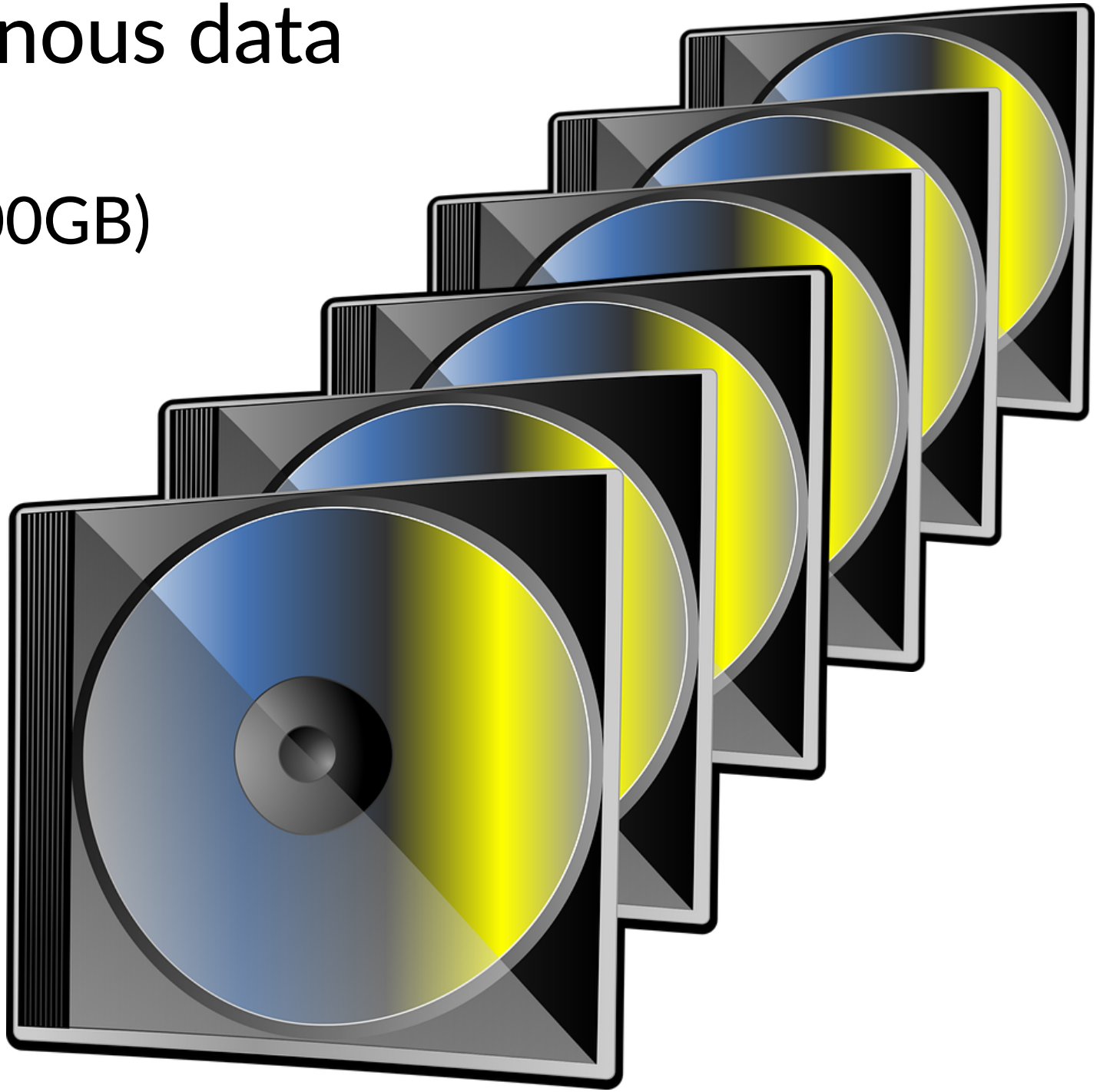
Keywords

Micro-computed tomography, systematics, taxonomy, 3D, visualisation, interactive PDF, polychaetes, cybertypes, cybertaxonomy

research data
(supplementary to paper)

Voluminous data

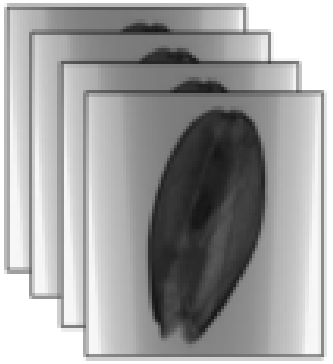
(up to 100GB)



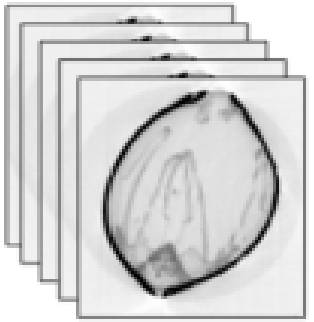
What to publish?

What to publish?

Raw data



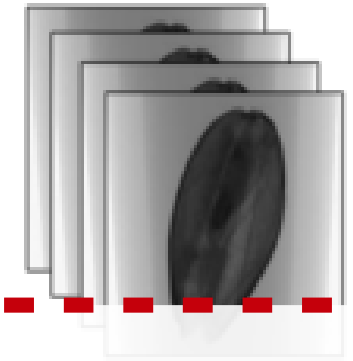
Processed data



Steps & analyses

What to publish?

Raw data



Processed data

+ metadata!

Steps & analyses

PROCEEDINGS B

rspb.royalsocietypublishing.org

Perspective

Cite this article: Davies TG *et al*. 2017 Open data and digital morphology. *Proc. R. Soc. B* **284**: 20170194.
http://dx.doi.org/10.1098/rspb.2017.0194

Received: 30 January 2017
Accepted: 10 March 2017

Subject Category:
Morphology and biomechanics

Open data and digital morphology

Thomas G. Davies[1], Imran A. Rahman[1,2], Stephan Lautenschlager[1,3], John A. Cunningham[1], Robert J. Asher[4], Paul M. Barrett[5], Karl T. Bates[6], Stefan Bengtson[7], Roger B. J. Benson[8], Doug M. Boyer[9], José Braga[10,11], Jen A. Bright[12,13], Leon P. A. M. Claessens[14], Philip G. Cox[15], Xi-Ping Dong[16], Alistair R. Evans[17], Peter L. Falkingham[18], Matt Friedman[19], Russell J. Garwood[5,20], Anjali Goswami[21], John R. Hutchinson[22], Nathan S. Jeffery[6], Zerina Johanson[5], Renaud Lebrun[23], Carlos Martínez-Pérez[1,24], Jesús Marugán-Lobón[25], Paul M. O'Higgins[15], Brian Metscher[26], Maëva Orliac[23], Timothy B. Rowe[27], Martin Rücklin[1,28], Marcelo R. Sánchez-Villagra[29], Neil H. Shubin[30], Selena Y. Smith[19], J. Matthias Starck[31], Chris Stringer[5], Adam P. Summers[32], Mark D. Sutton[33], Stig A. Walsh[34], Vera Weisbecker[35], Lawrence M. Witmer[36], Stephen Wroe[37], Zongjun Yin[1,38], Emily J. Rayfield[1] and Philip C. J. Donoghue[1]

[1]School of Earth Sciences, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ, UK
[2]Oxford University Museum of Natural History, Parks Road, Oxford OX1 3PW, UK
[3]School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham B15 2TT, UK
[4]Museum of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK

Where to publish?

Where to publish?

Not as supplementary data to paper

Where to publish?

Not as supplementary data to paper

Data repository

Where to publish?

molecular biology

NCBI

GENBANK

biogeography

GBIF

OBIS
OCEAN BIOGEOGRAPHIC INFORMATION SYSTEM

Where to publish?

- Dryad
- Figshare
- GigaDB
- Github
- Zenodo
- MorphoBank
- MorphoDBase
- MorphoMuseuM
- MorphoSource

Where to publish?

- Dryad
- Figshare
- GigaDB
- Github
- Zenodo
- MorphoBank
- MorphoDBase
- MorphoMuseuM
- MorphoSource

Type of data (volume, surface, analyses...) ?
Size restriction?
Costs?
Type of license for re-use?
Globally unique identifier (DOI) offered?

Data papers


(GIGA)n SCIENCE
Search GigaScience for Go
Advanced search
Home
Articles
Authors
Reviewers
About this journal
My GigaScience
GigaScience aims to revolutionize data dissemination, organization, understanding, and use. An online open-access open-data journal, we publish 'big-data' studies from the entire spectrum of life and biomedical sciences. To achieve our goals, the journal has a novel publication format: one that links standard manuscript publication with an extensive database that hosts all associated data and provides data analysis tools and cloud-computing resources.
Our scope covers not just 'omic' type data and the fields of high-throughput biology currently serviced by large public repositories, but also the growing range of more difficult-to-access data, such as imaging, neuroscience, ecology, cohort data, systems biology and other new types of large-scale sharable data.
Editorial Board | Editorial Team | Instructions for authors | FAQ
Get Credit for Your Data
Publish a Data Note of your research in GigaScience
Submit a manuscript
Register
Sign up for article alerts
Contact us
Follow us on Twitter
Find us on Facebook
Support
SCIENTIFIC DATA
Home | About | For Authors | For Referees | Advisory and Editorial Board | Open Access | FAQ
For Authors > Submission guidelines
Submission guidelines
The submission guidelines contain detailed information to help authors prepare, format and submit a Data Descriptor manuscript. Please see our guide to authors for additional information and policies relevant to authors, and view our first publications.
Scope guidelines
Preparing and submitting a Data Descriptor manuscript
Templates
Cover letter
Data deposition
Format of Data Descriptor manuscripts
Submitting experimental metadata
Statistical guidelines
Chemical and biological nomenclature and abbreviations
Gene nomenclature
Characterization of chemical and biomolecular materials
References
Equations
Scientific Data is accepting submissions
Submit manuscript
Scientific Data Updates
Scientific Data now inviting submissions from social sciences
November 13, 2014
Data Matters: Interview with Tony Hey
November 4, 2014
Data Matters: Interview with Anne Guex
October 28, 2014
Biodiversity Data Journal
Making your data count!
PENSOFT

No metadata standards exist

Multitude of data repositories

Publication = time consuming & difficult due to data volume

No metadata standards exist

Multitude of data repositories

Publication = time consuming & difficult due to data volume

Many datasets not published
Online datasets not retrievable

No metadata standards exist

Multitude of data repositories

Publication = time consuming & difficult due to data volume

Many datasets not published
Online datasets not retrievable

Common registry / metadata repository