

Sustainability in Biodiversity Software Development: More financing or better practices?

**Matthew Yoder,
Michael Twidale,
Andrea Thomer**

TDWG 2017, Ottawa

<https://goo.gl/KHLekS>

<https://github.com/mjy/presentations>

```
if title =~ /.+:.+or.+?$/ then
```

Yes

Acknowledgements

Dmitry Mozzherin

Species File Group

Jim Beach

NSF-ABI-1356381

Atom Dark Theme

> |

Overview

Species File Group (SFG)

How did this happen?

What have we been doing?

How might we evolve?

Software sustainability

Financing

Practices

Visualizing sustainability

Species

File

Group

Initial funding of software

Private (EMu)

NSF (Arbor, Open Tree, Specify)

Institutional (Scratchpads)

Research by-product (Phylogenetic analyses)

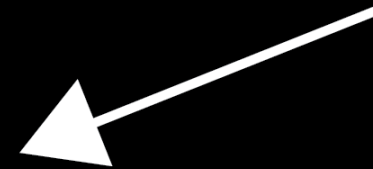
National/International Organization (GBIF, ALA)

Event-based (Hackathons, Workshops)

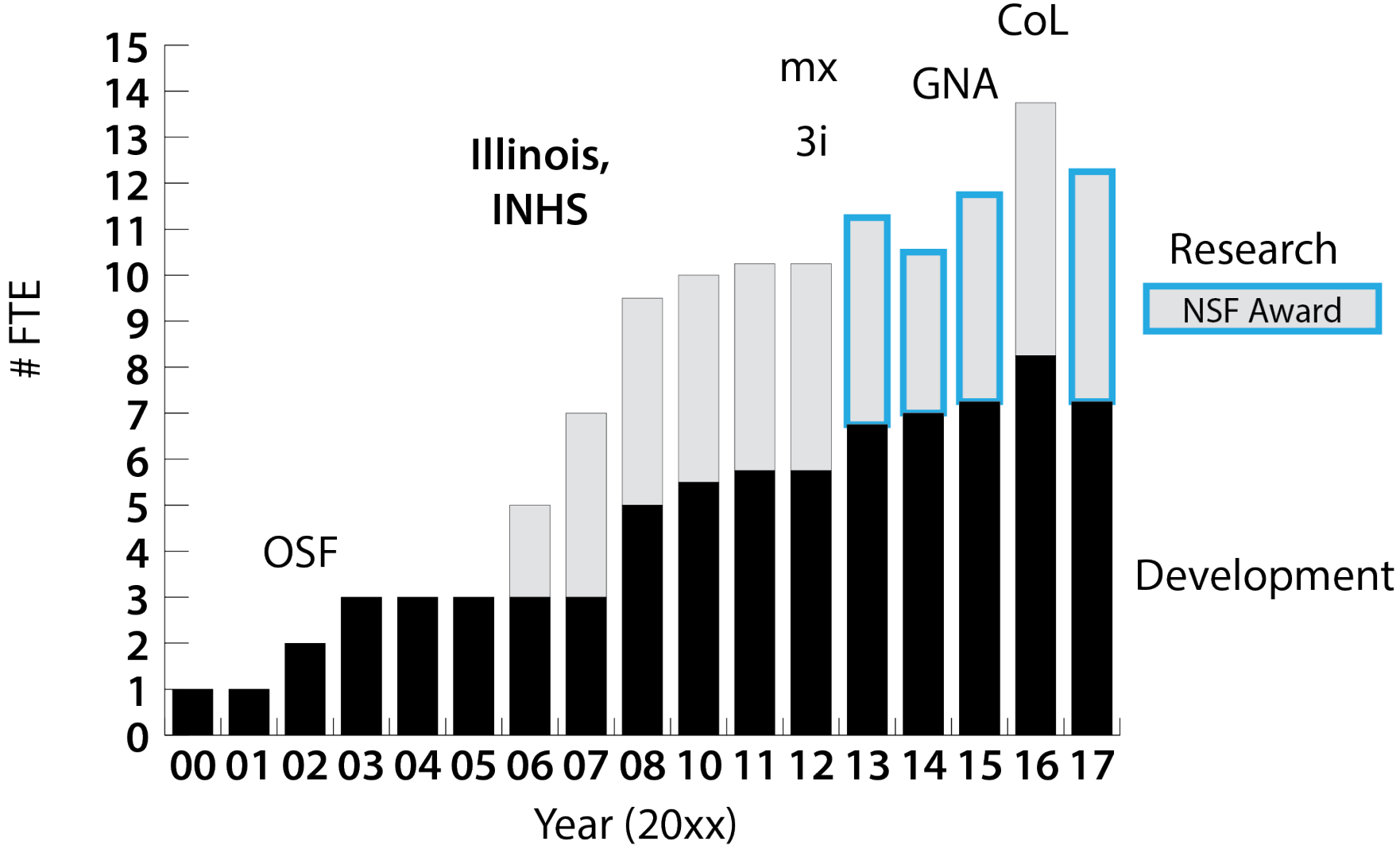
Philanthropy (EOL)

Participatory Philanthropy (Species File Group)

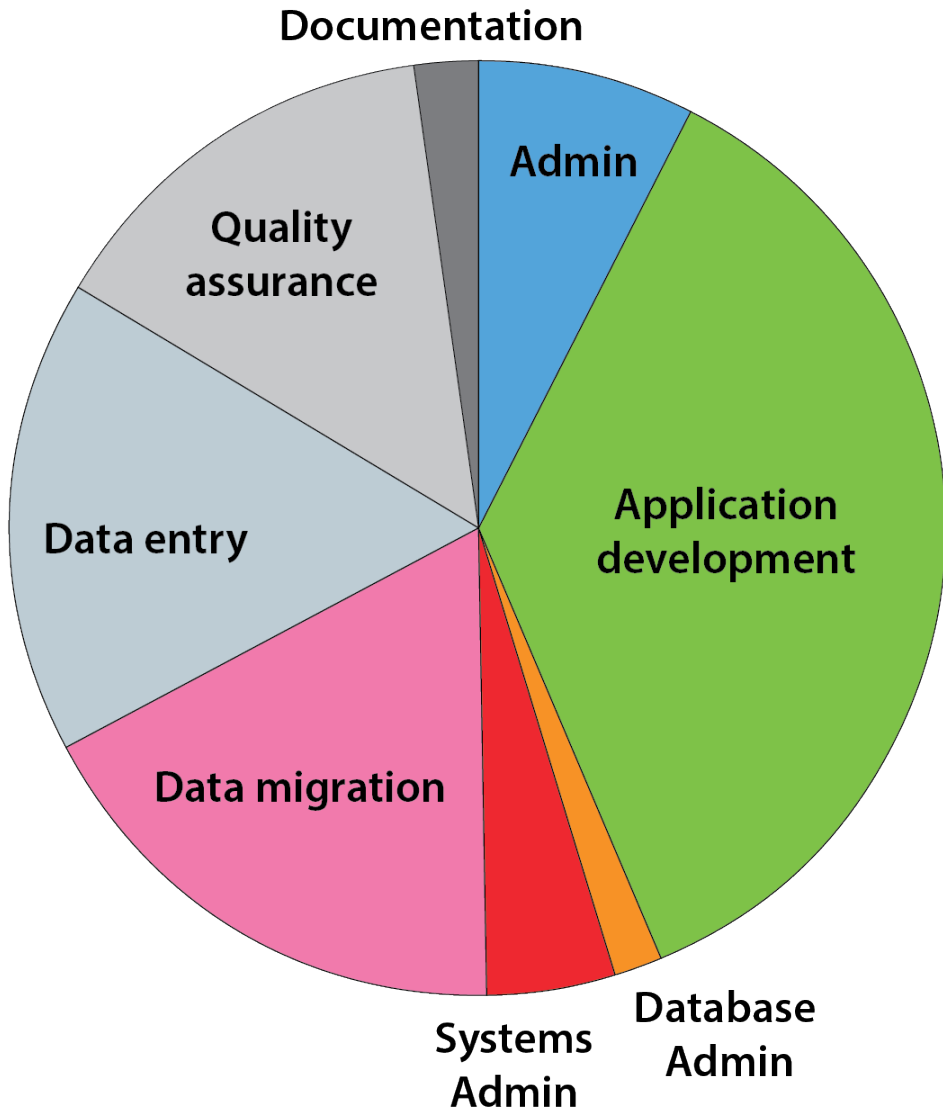
Collaboration (Symbiota, Arctos)



SFG History



SFG - Informatics effort breakdown*



*payroll only

Workbenches

Treehoppers

Aetalionidae, Melizoderidae, and Membracidae (Hemiptera)



- Introduction
- Search Taxa
- Higher Taxa
- Genera
- Keys
- Top Pics
- Sounds
- Distribution
- References
- Links
- News

Introduction

TREEHOPPERS
biblio
infor
Meliz
treeh
shap
quest
webs
amor
publ
Overview
Treeh
treeh
gene
linea
zooe
speci
A...

3I Interactive Keys and Taxonomic Databases

Dmitry A. Dmitriev



3I (Internet-accessible Interactive Identification) is a set of software tools for creating on-line identification keys, taxonomic databases, and virtual taxonomic revisions. By organizing illustrations and

Best viewed with IE 5.0+

Hymenoptera Anatomy Ontology Portal



go: search | analyze | give feedback | references | terms | tree | partonomy | pulse | about / how to cite

Search the Ontology

Search Show



Orthoptera Species File (Version 5.0/5.0)

Home Search Taxa Key Help Wiki

Orthoptera Species File Online

This HAO portal is a project of the Hy in part by NSF grants BDI-0446224, i and not the NSF.

- María Marta Cigliano**, Author,
División Entomología, Museo de La Plata, CEPAVE - CONICET (National Scientific and Technical Research Council), Argentina
- Holger Braun**, Author,
División Entomología, Museo de La Plata, Argentina
- David C. Eades**, Principal Database Developer,
Illinois Natural History Survey
- Daniel Otte**, Founder and Principal Author,
Academy of Natural Sciences of Philadelphia

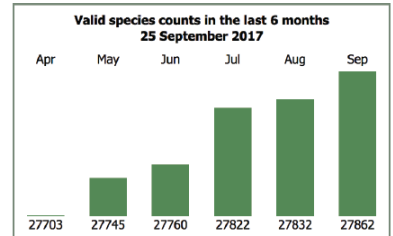
- Major Contributors:**
- Piotr Naskrecki**, Museum of Comparative Zoology, Harvard University
- Hernán Lucas Pereira**, CEPAVE - CONICET (National Scientific and Technical Research Council), Argentina

List of experts

With the cooperation of
The Orthopterists' Society



Copiphora rhinoceros Pictet, 1888



[Vision](#)

[Features](#)

[Getting started](#)



TaxonWorks

Describe life.

TaxonWorks is an integrated web-based workbench for taxonomists and biodiversity scientists.

It allows you to capture, organize, and enrich your data; share it with collaborators; and package it for analysis and publication.

Global Names Architecture

[HOME](#)[NEWS](#)[APPS](#)[PAPERS](#)[DOCUMENTATION](#)[SOURCE](#)

Global Names Architecture

Scientific Names Services

The Global Names Architecture (GNA) is a system of web-services which helps people to register, find, index, check and organize biological scientific names and interconnect on-line information about species.

Resolve

Check if the names you have are real, spelled right, and currently used.

[Resolver](#) →

Find

Discover scientific names in texts, PDFs, and even images you took in a Natural History Museum.

[Name Finder](#) →

Register

If you are a zoologist -- you can officially register species electronically.

[Zoobank](#) →



The content of this website is in public domain

Powered by [Jekyll](#)

Hosted on [GitHub](#)

Catalogue of Life



The Importance Of Global Species Databases

Find out why the Catalogue of Life is the most authoritative global species index available.

[Read More](#)



Welcome to the Catalogue of Life website. The gateway to our online database of the world's known species of animals, plants, fungi and micro-organisms.

There are two distinct versions of the Catalogue: the Catalogue of Life monthly edition and the Catalogue of Life Annual Checklist. Choose the version most suited to your needs.

[Read More](#)

FAQ

What is the Catalogue of Life?

The Catalogue of Life is the most comprehensive and authoritative global index of species currently available. It consists of a single integrated species checklist and taxonomic hierarchy. The Catalogue holds essential information on the names, relationships and distributions of over 1.6 million species. This figure continues to rise as information is compiled from diverse sources around the world.

[Read more](#)

Catalogue of Life: August 2017

Monthly edition

Next update: September 2017

[Access](#)

Catalogue of Life: 2017 Annual Checklist

A referenceable snapshot once per year

Next publication date: April 2018

[Access](#)

Progress

Total Species: > 1,6 M

Total contributing databases: 156

84%

Orthopterists' Society

Contact us



THE ORTHOPTERISTS' SOCIETY

HOME

News archive

Membership ▶

Publications ▶

Resources ▶

Bulletin board

About us



Welcome to the Orthopterists' Society website

The Orthopterists' Society (formerly the Pan American Acridological Society) is an international scientific organization devoted to facilitating communication among those interested in Orthoptera and related organisms. We foster research and publication in all aspects of the biology of these insects from ecology and taxonomy to physiology, endocrinology, cytogenetics, and control measures.

Latest news

09/20/2017

New *Metaleptea*



The latest issue of *Metaleptea* (37/3) is now available for download.

[Download the latest issue \(37/3\)](#)

[Access all issues of *Metaleptea*](#)

Posted by Webmaster

09/20/2017

Journal of Orthoptera Research is now open access



JOR

As of 1st of May, 2017, the Orthopterists' Society has signed a publication agreement with Pensoft. JOR is now Open Access, which means that any member of the public can freely download articles, increasing JOR's accessibility to a wider audience. For members of the Orthopterists' Society, publication in JOR is free of charge. Online submission and review are managed through Pensoft's state-of-the-art ARPHA system and articles appear online as soon as they are accepted for publication.

[JOR at Pensoft](#)

[Access all issues of JOR](#)

Posted by Piotr Naskrecki

05/18/2017

New *Metaleptea*



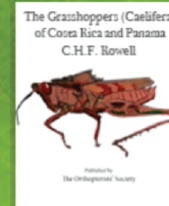
The latest issue of *Metaleptea* (37/2) is now available for download.

[Download the latest issue \(37/2\)](#)



Download a PDF brochure

About the Orthopterists' Society



OUT NOW!

SFG Research

Grants (program, SFG PI)

ABI - Yoder - Taxonomic Interfaces, Anatomy Ontologies

ABI - Mozzherin - Global names

CSBR - DeWalt, Yoder, Dmitriev - Orphaned collections

DEB - Yoder - Chalcidoid Phylogenetics

DEB - Dmitriev - GoLife on planthopper phylogenomics

Species File Group Evolution

support and advance biodiversity informatics

Application/service developers

Github: 2 organizations, 75 repositories

Hosting

~60 active projects globally

Data migrators

Research

Biodiversity Informatics Hub

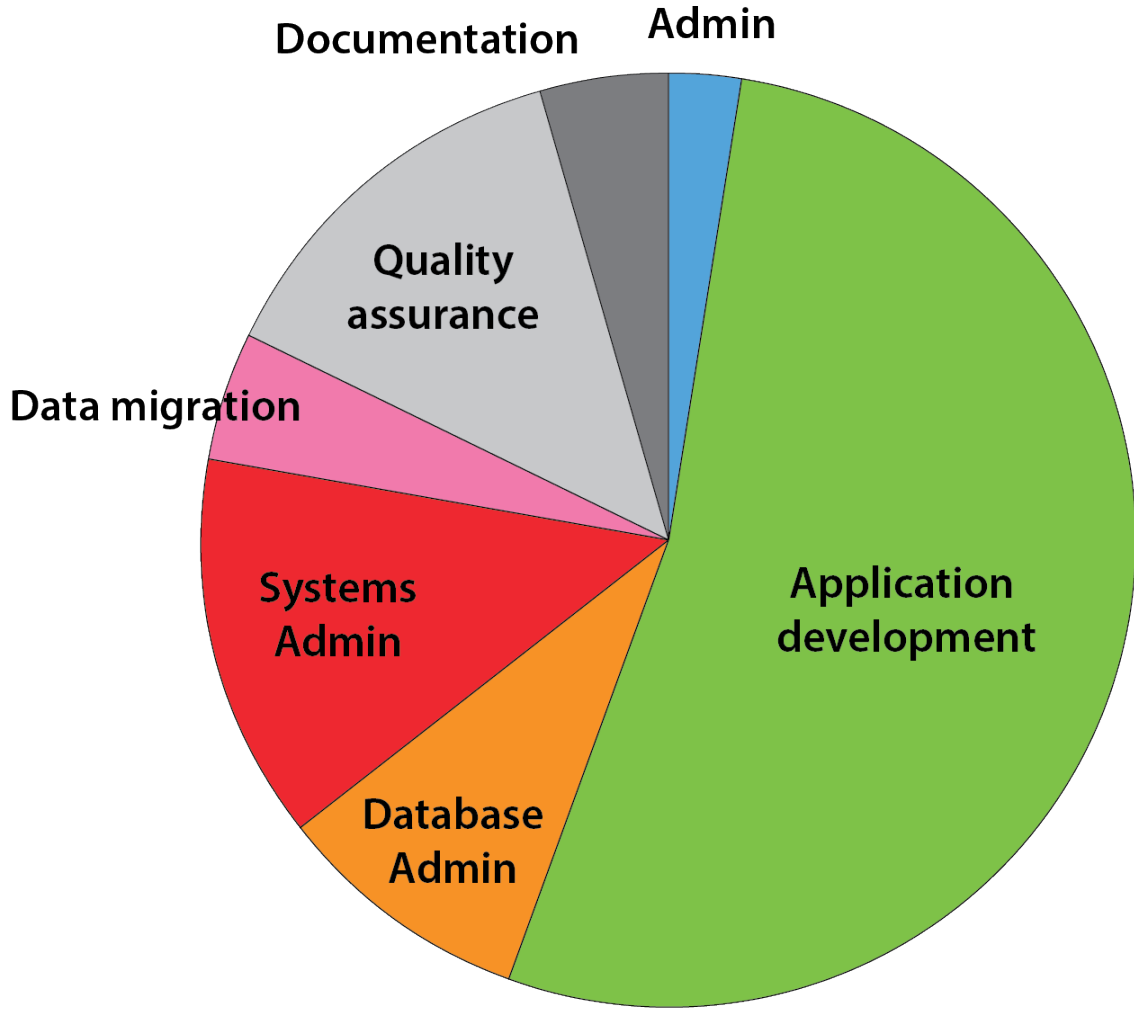
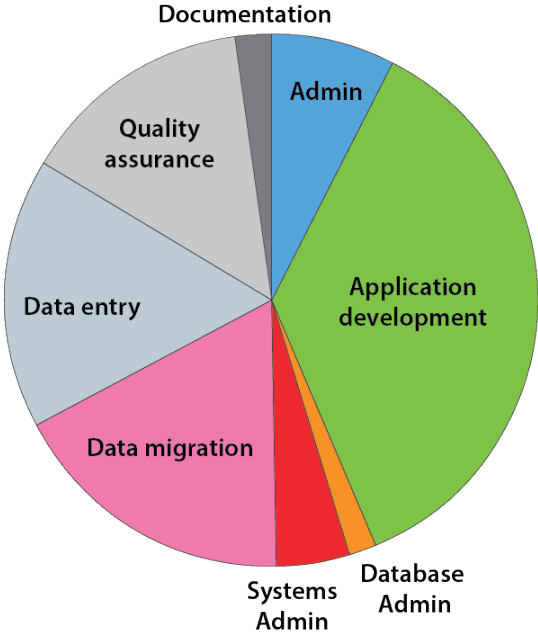
Education

Short term visits

Hackathons

Training

If SFG goal is to improve software sustainability



Integrative Biodiversity Informatics Team

Back-end developer

“Compute” developer (AI/Machine learning etc.)

Data archivist (both migrations and archiving)

Database administrator / Data scientist

Systems administrator

2 User interface developers

User support / Information Scientist

User experience (UX) / Information scientist

3-4 Research programmers (Biodiversity Informaticians)

Rotating resident biologist (non programmer)

Software

Sustainability

Sustainable funding of software

Private (EMu)	Expensive, uncertain
NSF (Arbor, Open Tree, Specify)	Innovation != sustaining
Institutional (Scratchpads)	Needs of the many; Revolving door
Research by-product (Phylogenetic analyses)	Temporary
National/International Organization (GBIF, ALA)	High barrier to entry
Event-based (Hackathons, Workshops)	Spontaneous
Philanthropy (EOL)	Temporary if not institutionalized
Participatory Philanthropy (Species File Group)	Scalability
Collaboration (Symbiota, Arctos)	Scalability

More funding collaboration

Scientific developer

Systems administrators

Database administrators

Semantics

User interfaces

Archivists

Design

Human-computer Interactions

User experience

More funding collaboration

Scientific developer

Systems administrators

Database administrators

Semantics

User interfaces

Archivists

Design

Human-computer Interactions

User experience

Add people via alternate entry points

more people, more sustainability

“Here is the user-manual”

or

“Code in language X”

**Play in a
sandbox**

Suggest improvements

Suggest a workflow

Help write the user manual

Draw an interface

File issues

Take a carpentry course

Participate in a hackathon

Code a feature

Practices

1 - Promote sustainability

**My code is open, here
it is on Github**

**Use other's
code first**

Code for others first

WRITE TESTS

Containerize environments

Continuously integrate

Reference "Gold standards"

Expose an API

Atomize libraries

Scaffold code

Practices

2- Lower the barrier to entry

**My code is open, here
it is on Github**

**Use other's
code first**

Code for others first

WRITE TESTS

Containerize environments

Continuously integrate

Reference "Gold standards"

Expose an API

Atomize libraries

Scaffold code

Practices

3 - Designing for “apoptosis”

**My code is open, here
it is on Github**

**Use other’s
code first**

Code for others first

WRITE TESTS

Containerize environments

Continuously integrate

Reference “Gold standards”

Expose an API

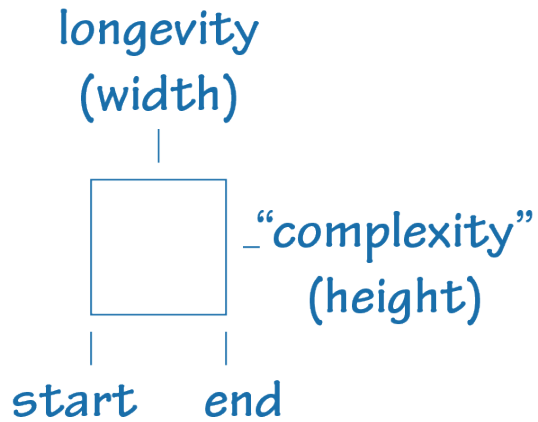
Atomize libraries

Scaffold code

**It is open
source,
right?**

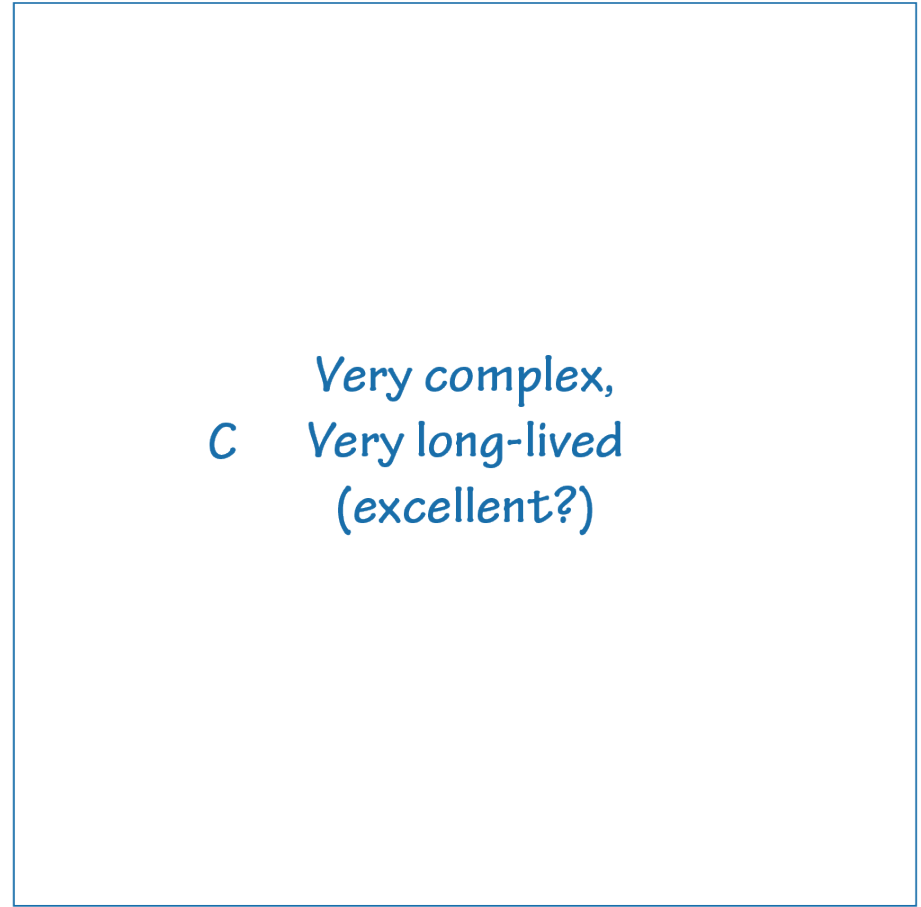
Open source in 2017, LUL

Blueprints for sustainable software



Very complex,
short-lived
(bad)

A



C Very complex,
Very long-lived
(excellent?)

B

Very simple, very long-lived
(good)

t_0

$t_0 + 10 \text{ years}$

Time

Blueprints
for
sustainable
software

Analysis 

Script 

Library 

Widget

Workbench 

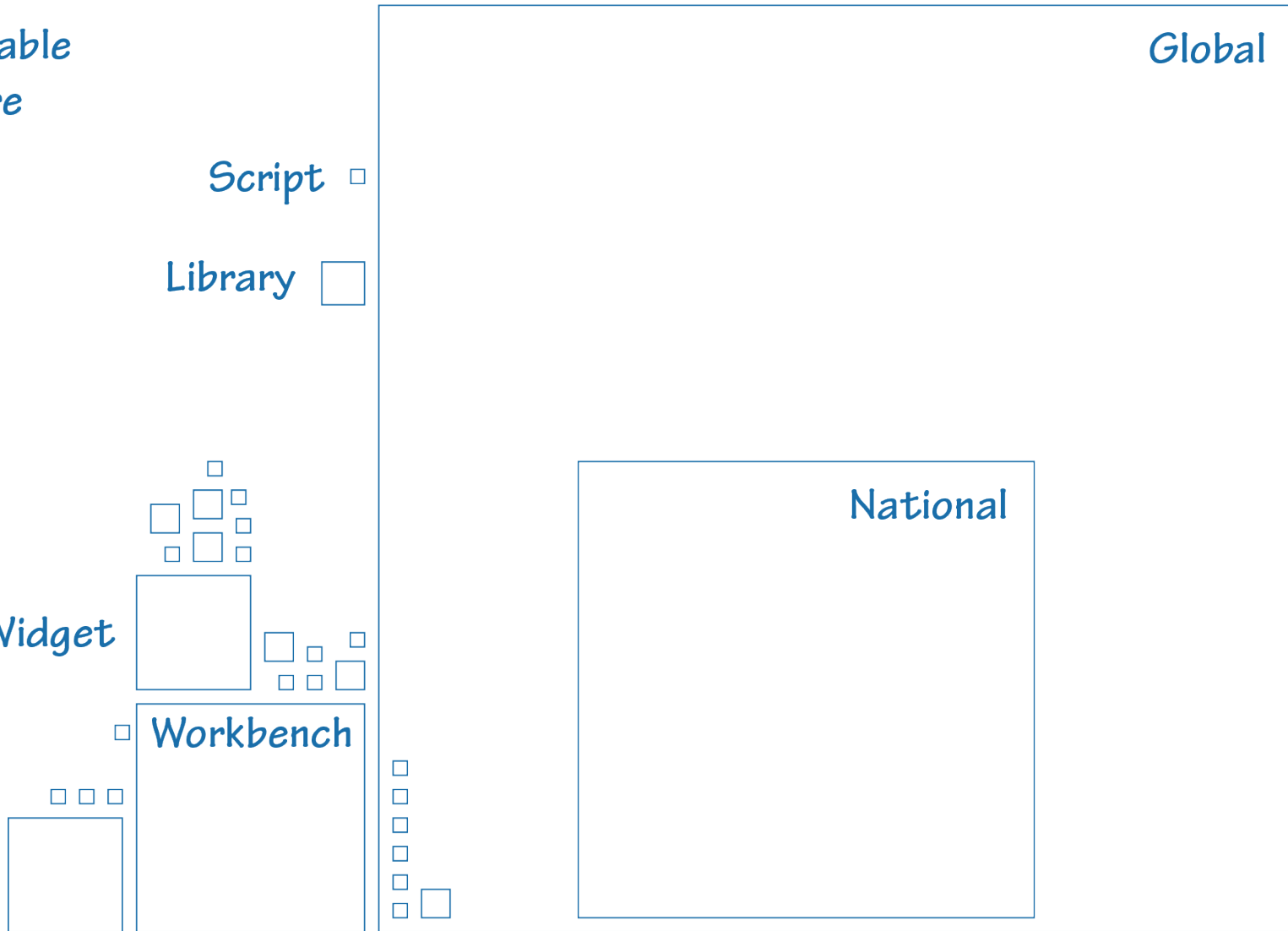
Global

National

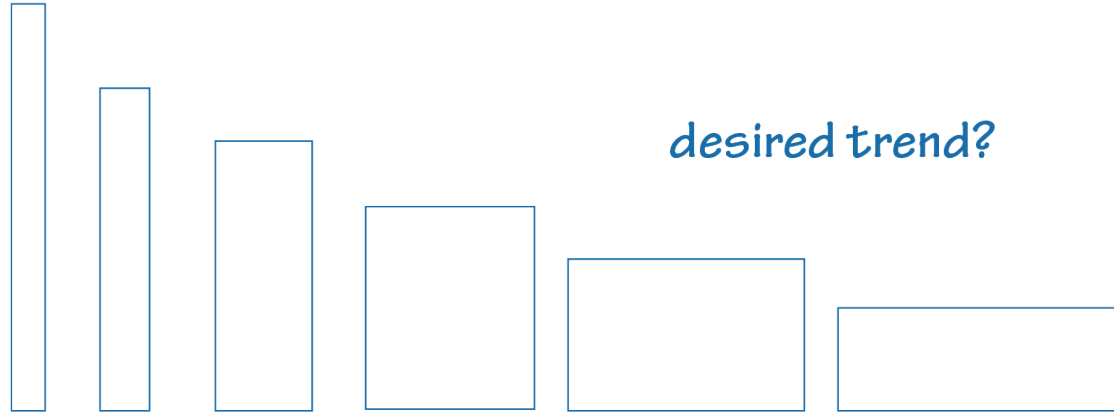
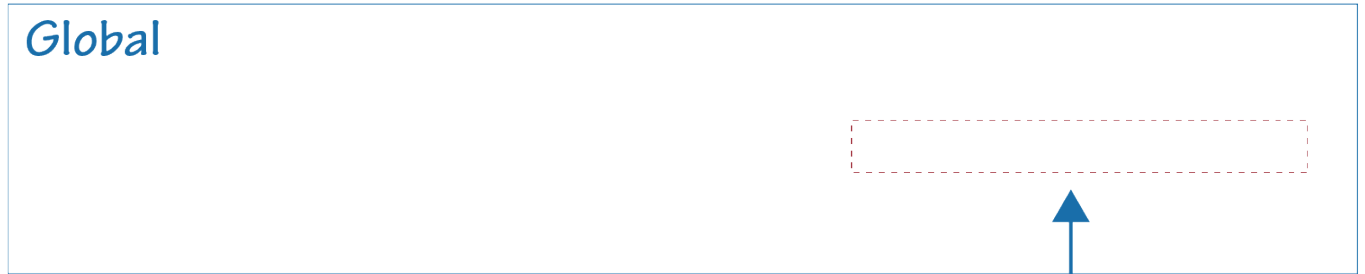
t0

t0+10 years

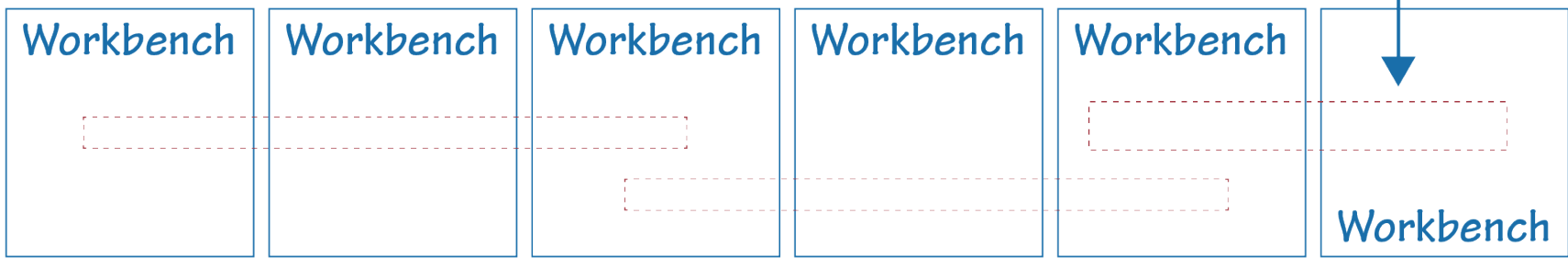
Time



Blueprints
for
sustainable
software



obsolescence is inevitable



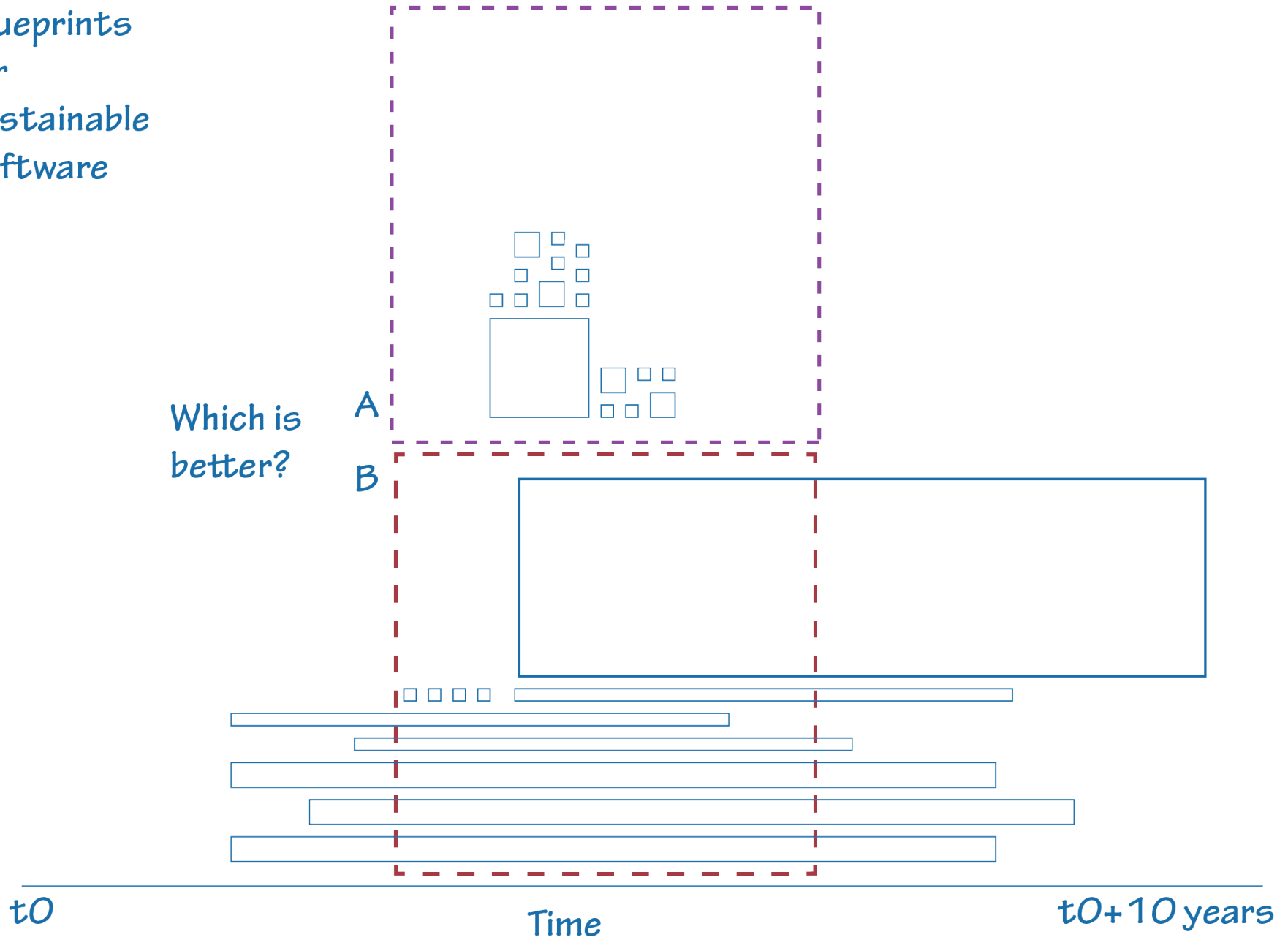
t0

t0+30 years

Time

Blueprints for sustainable software

Which is
better?



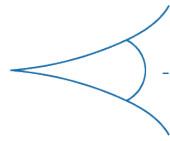
Blueprints for sustainable software

Software rarely implements the full standard

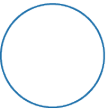
Interface



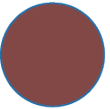
Selection Pressure



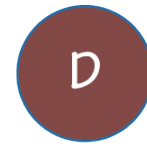
Standard



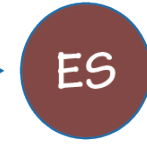
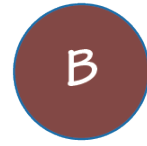
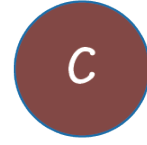
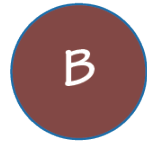
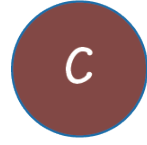
TDWG



Bespoke



Removed from "population"



Added to standard

Interface as standard?

t0

t1

t2

Time



Thank you

Links

<https://github.com/SpeciesFileGroup/>
<https://github.com/GlobalNamesArchitecture>
<http://taxonworks.org>
<https://twitter.com/taxonworks>
<http://orthoptera.speciesfile.org>
<http://treehoppers.insectmuseum.org>
<http://dmitriev.speciesfile.org/>
<http://orthsoc.org/>
<http://catalogueoflife.org/>
<http://portal.hymao.org>
<http://tenebrionidbase.org>

Acronyms

EOL - Encyclopedia of Life
mx - <http://purl.oclc.org/NET/mx-database>
AI - Artificial Intelligence
ML - Machine Learning
SFG - Species File Group
TDWG - Taxonomic Databases Working Group
FTE - Full time employee
INHS - Illinois Natural History Survey
OSF - Orthoptera Species File
GBIF - Global Biodiversity Information Facility
ALA - Atlas of Living Australia
LUL = LOL - Laughing Out Loud. Gamer meme.
In this case, roughly:
“so obvious now it shouldn’t be raised, but it is anyways, and that’s a little silly”