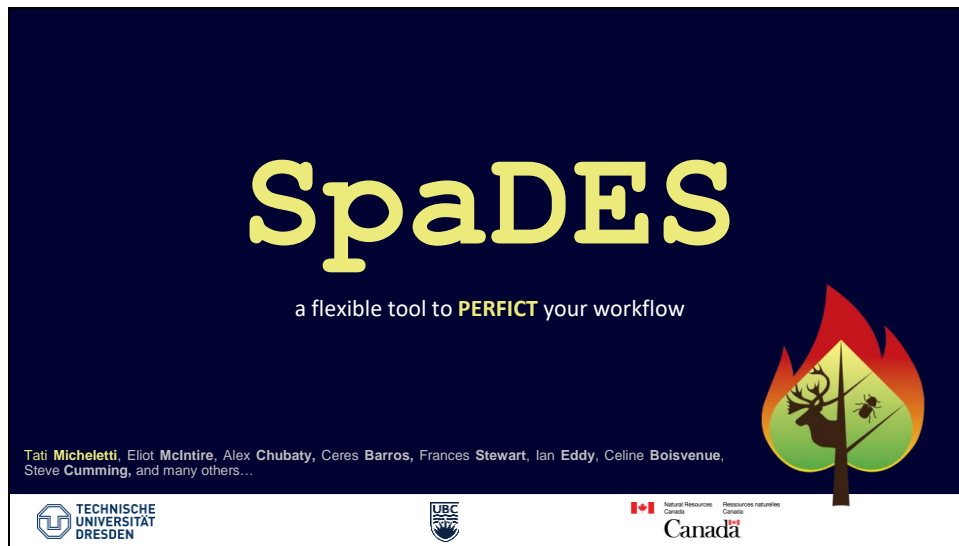
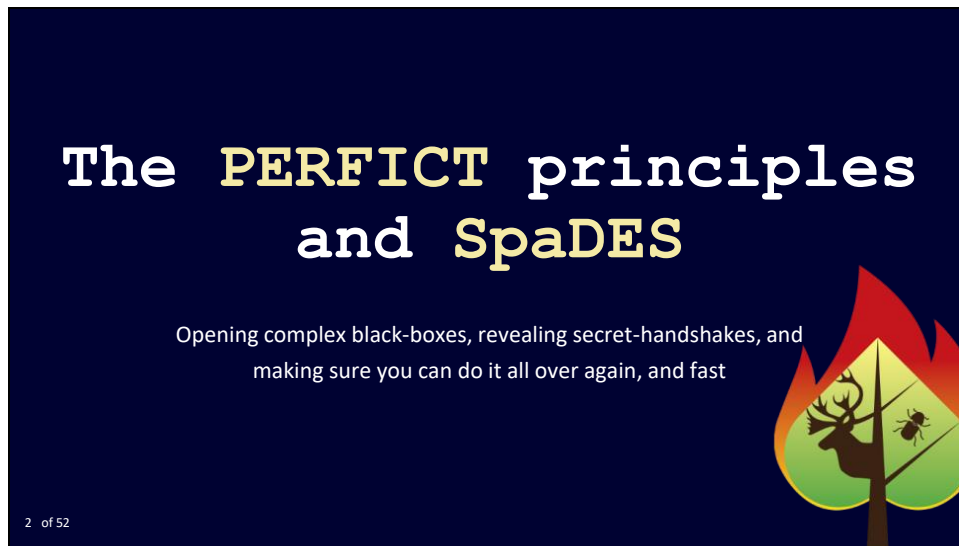


Slide 1



Hi, thank you everyone for being here, and thank you EFI for the invitation. My name is Tati Micheletti, I am a research associate at the University of British Columbia Canada, but starting next month at Dresden University of Technology in Germany. I will tell you today about SpaDES.

Slide 2



I would like to start with a short story.

About 20 years ago, there was an ecologist who, together with other scientists, dreamed of a modelling tool that could do it all: facilitate frequent predictions, perform model evaluations and validation, a tool that would make models not only reproducible, but reusable, a tool that would promote models to be free and accessible, that would allow for interoperability among models, and which would support continuous workflows, that could be routinely tested. As he couldn't find any tools that would check all the boxes, he decided to write one! He partnered with a postdoc and they started to work on what would become SpaDES. In August 2017, SpaDES made its way to CRAN.

Slide 3

The image shows the front cover of a Scientific Data journal article. At the top right is the URL 'www.nature.com/scientificdata'. The title 'SCIENTIFIC DATA' is prominently displayed in a large, sans-serif font, with a decorative graphic of binary code (0s and 1s) to its right. Below the title, the text 'Amended: Addendum' is visible. The article title 'Comment: The FAIR Guiding Principles for scientific data management and stewardship' is written in a bold, blue font. The authors 'Mark D. Wilkinson et al.*' are listed below the title. On the left side, under the heading 'SUBJECT CATEGORIES', there are two sub-points: '» Research data' and '» Publication characteristics'. At the bottom left, the publication timeline is given: 'Received: 10 December 2015', 'Accepted: 12 February 2016', and 'Published: 15 March 2016'. A small page number '3 of 52' is located in the bottom left corner of the slide frame.

www.nature.com/scientificdata

SCIENTIFIC DATA

Amended: Addendum

Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson et al.*

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measurable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. This Comment is the first formal publication of the FAIR Principles, and includes the rationale behind them, and some exemplar implementations in the community.

SUBJECT CATEGORIES

- » Research data
- » Publication characteristics

Received: 10 December 2015
Accepted: 12 February 2016
Published: 15 March 2016

3 of 52

Before talking about PERFICT, we will talk about FAIR. The FAIR principle was proposed by several scientists in 2016 and stands for Findable, Accessible, Interoperable and Reusable is a concept defined by several scientists in a workshop done in 2014 regarding data standards. Since then, this guideline has also been suggested for models. PERFICT is an expansion of the FAIR data principles and models to modelling workflows. PERFICT stands for Making Predictions, evaluate models, make models Reusable, freely accessible and Interoperable, built within Continuous workflows that are routinely Tested.

Slide 4

The PERFICT principle & Spades

PERFICT

Make Predictions, Evaluate models, make models Reusable, Freely accessible and Interoperable, built within Continuous workflows that are routinely

4 of 52

ECOLOGY LETTERS

VIEWPOINT | Open Access | CC BY

PERFICT: A Re-imagined foundation for predictive ecology

Ellor J. B. McIntire, Alex M. Chubaty, Steven G. Cumming, Dave Anderson, Ceres Barros, Celine Boisvenue, Samuel Haché, Yong Luo, Tatiane Micheletti, Frances E. C. Stewart

First published: 22 March 2022 | <https://doi.org/10.1111/ele.13994> | Citations: 9

SECTIONS | PDF | COMMENTS | TOOLS | SHARE

Abstract

Making predictions from ecological models—and comparing them to data—offers a coherent approach to evaluate model quality, regardless of model complexity or modelling paradigm. To date, our ability to use predictions for developing, validating, updating, integrating and applying models across scientific disciplines while influencing management decisions, policies, and the public has been hampered by disparate perspectives on prediction and inadequately integrated approaches. We present an updated foundation for Predictive Ecology based on seven principles applied to ecological modelling: make frequent Predictions, Evaluate models, make models Reusable, Freely accessible and Interoperable, built within Continuous workflows that are routinely Tested (PERFICT). We outline some benefits of working with these principles: accelerating science; linking with data science; and improving science-policy integration.

PERFICT was defined on the 2022 Ecology Letter's paper from McIntire and collaborators and is an appeal to the scientific community to accelerate the creation and continued use of shared knowledge.

Slide 5

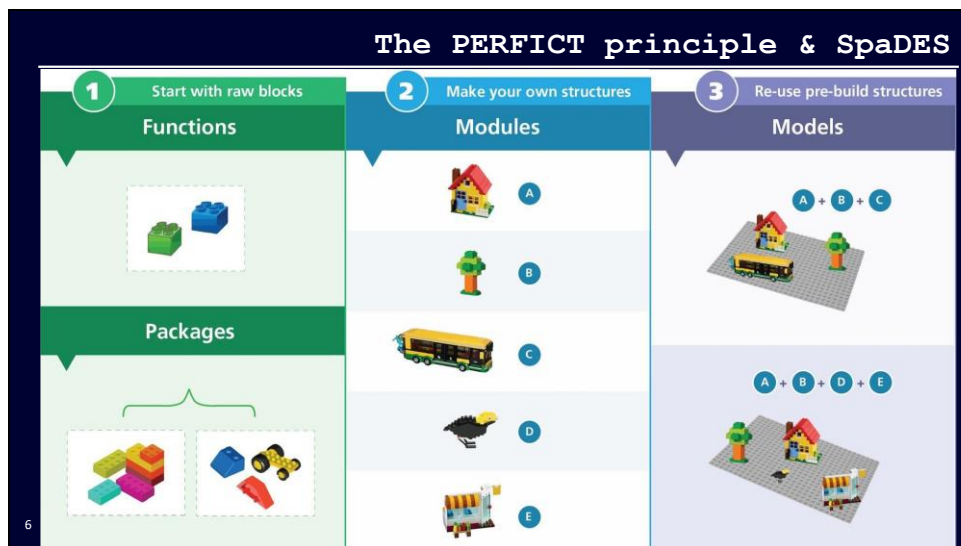
The PERFICT principle & SpaDES

- ♠ **PERFICT**: FAIR Principles brought to a modelling workflow context
- ♠ Foundational principles for which **SpaDES** was built

5 of 52

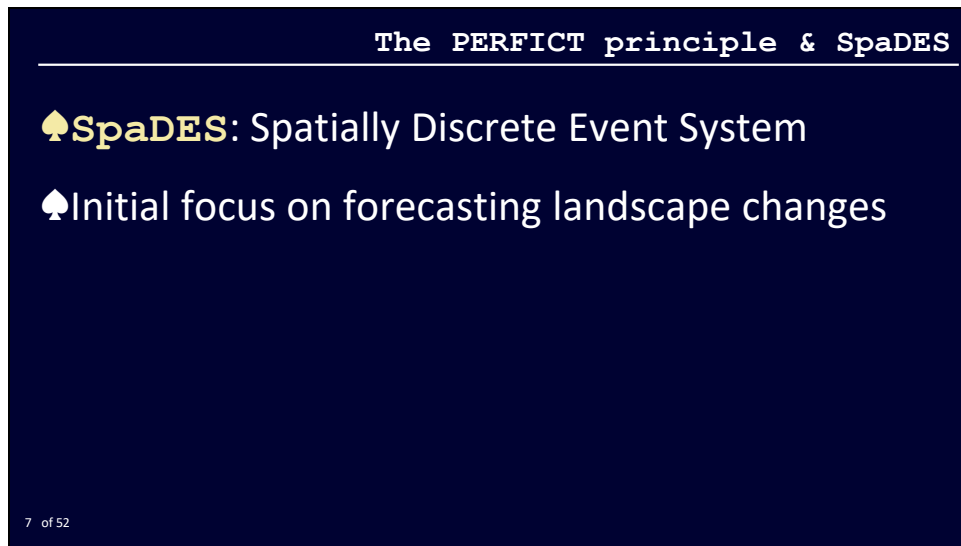
So in a nutshell, PERFICT are the FAIR principles brought to a modelling workflow context and the foundational principles for which SpaDES was built. SpaDES is the tool to help implement these such principles in modelling workflows.

Slide 6



Here we can see an abstraction of the PERFICT principles. Starting from raw blocks (which can be seen as functions in R), we can build Modules, which convey specific processes (for example, “Data Acquisition and Treatment”, or “Parameterizing and Fitting a Species Model”). These modules can be combined into “Models” (in the wide sense of the word), which aim at answering specific questions. SpaDES is the tool that makes it possible to reuse and recombined these prebuilt structures (a.k.a. the interoperability and reusability points in FAIR and PERFICT).

Slide 7



The PERFICT principle & SpaDES

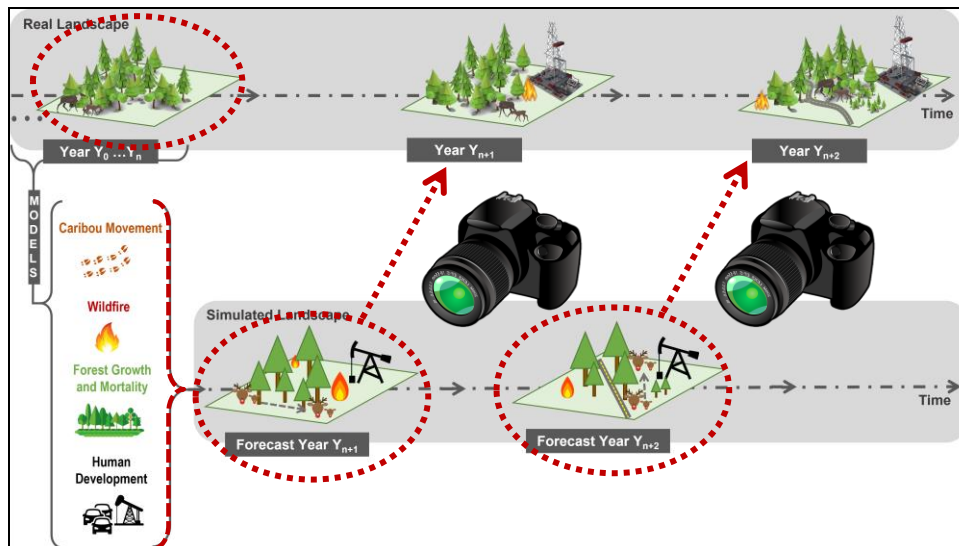
♠SpaDES: Spatially Discrete Event System

♠Initial focus on forecasting landscape changes

7 of 52

So, what is SpaDES? As our short story explained, SpaDES is a tool that allows researchers to PERFICT their workflows. SpaDES stands for Spatially Discrete Event System, which means that each “step” or “event” happens as a snapshot. It was originally designed with focus on forecasting landscape changes.

Slide 8



For example, when we observe and collect data from landscape, we can build models for its processes. In this example, a caribou movement, a wildfire, a forest growth and mortality, and a human development. These models can then in turn be used to simulate into the future (known forecasting), the landscape at a future discrete point in time, like a snapshot.

The PERFICT principle & SpaDES

- ♠ **SpaDES**: Spatially Discrete Event System
- ♠ Initial focus on forecasting landscape changes
- ♠ But... SpaDES is useful for full workflows
- ♠ Agnostic to modelling approach
- ♠ It enables the implementation of the **PERFICT** principles to workflows from the start

9 of 52

But a dedicated simulation platform only does a small part of the project, arguably, the simplest. We needed a platform that could deal with the whole workflow, from data acquisition, to treatment, to parameterization, to model fitting, to predicting and forecasting, to model validation, to posthoc analysis of results, to building figures and plots, to benchmarking, to supporting decisions. So that's what SpaDES became: a tool for PERFICT-ing workflows that is agnostic to the modelling approach.

The PERFICT principle & SpaDES

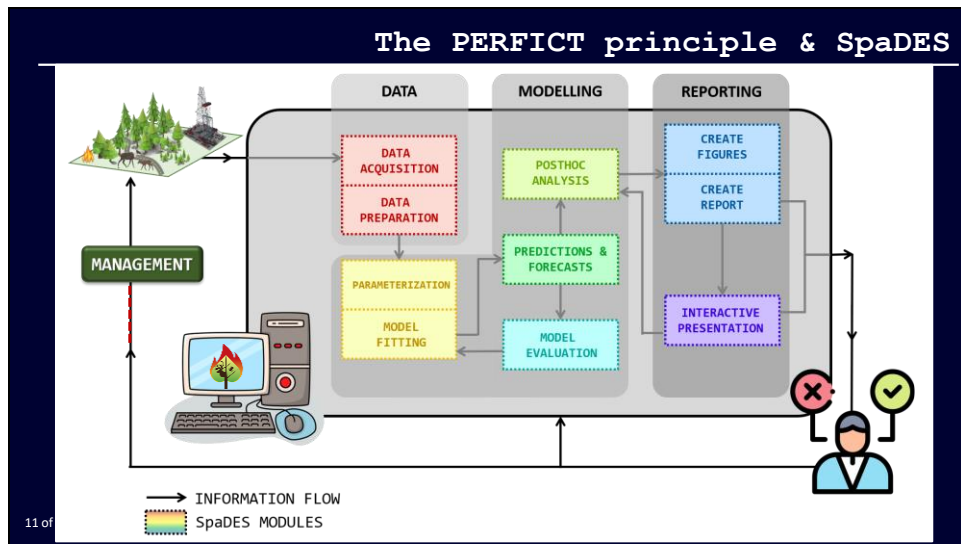
- ♠ **SpaDES** is a tool for developing and sharing modular code based on the **PERFICT** principles
- ♠ **SpaDES** supports the creation of SpaDES compatible **modules** and orchestrate their interactions to PERFICT workflows

Central piece!

10 of 52


Specifically, SpaDES is a tool for developing and sharing modular code based on the PERFICT principles. It supports the creation of SpaDES compatible modules and orchestrate their interactions to PERFICT workflows. The modular part of SpaDES comes from what we call **SpaDES modules**, which are SpaDES' central piece.

Slide 11



These modules represent stand-alone processes in the workflow (the dashed rectangles), such as “Acquire and Prepare Data”, “Parameterize and Fit the Model”, “Evaluate the Model”, “Make Predictions and Forecasts” , “Perform Posthoc Analysis”, “Prepare Figures and Reports”, “Present Results interactively”... and the hope is that this streamlined workflow can help decision making and get feedback from it to be iteratively improved.

The PERFICT principle & SpaDES




Metapackage:

- reproducible
- SpaDES.core

Other packages from the PERFICT Universe:

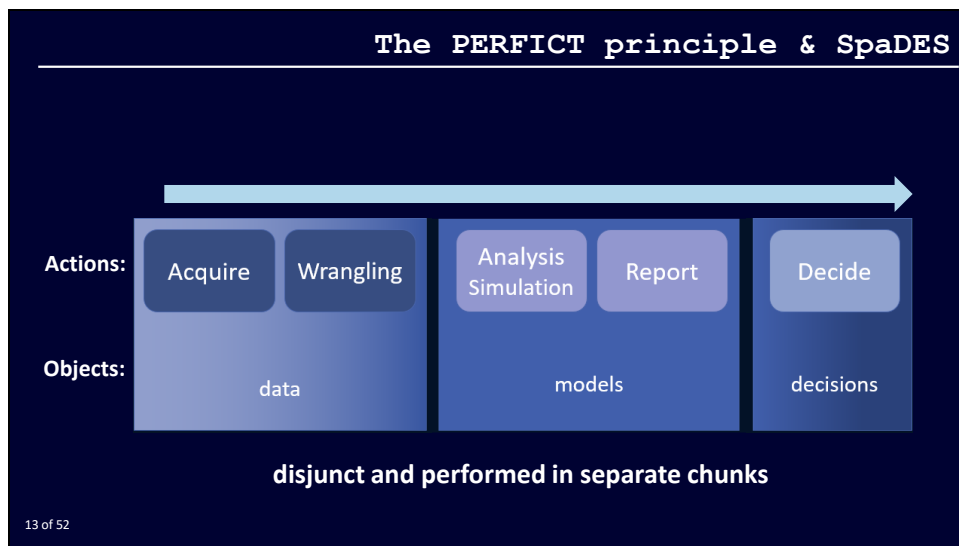
- SpaDES.tools
- SpaDES.project
- SpaDES.experiment
- SpaDES.shiny
- quickPlot



12 of 52

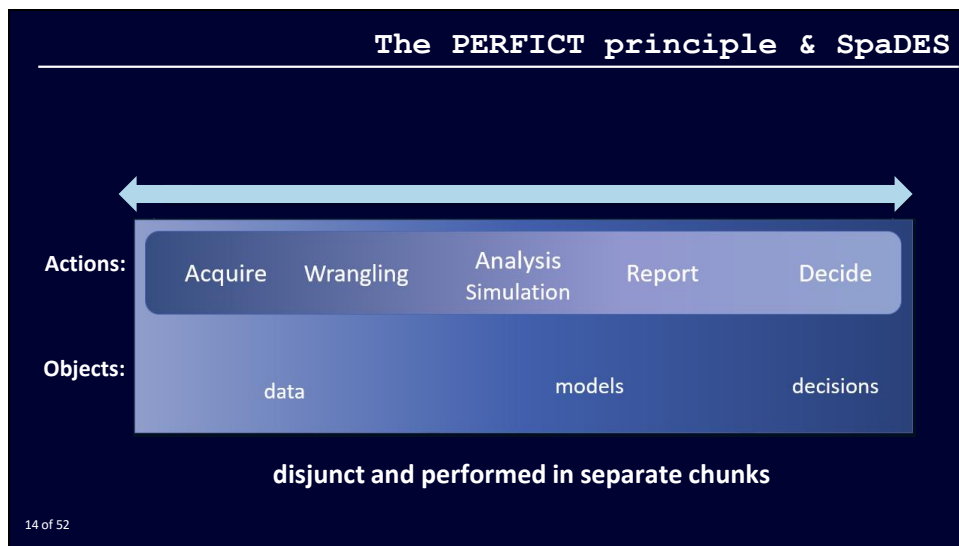
Technically speaking, SpaDES is an R metapackage, composed mainly by two packages: **reproducible** and **SpaDES.core**. These are part of a larger collection, the **PERFICT Universe**. Other packages from this collection include functionalities to help **run replicates** of simulation experiments, **create shiny apps**, provide **NetLogo and SELES functionalities**, provide **efficient plotting functions**, etc.

Slide 13



In a real-life context, if we think on the whole process of using models for decision making, most likely it will look like something like this: first step, acquiring and wrangling data, second step, performing analysis and/or simulations, and producing a report, last step, getting the report to the hands of decision makers. The problem with this workflow, is that it is disjunct, performed in chunks, likely performed by different teams, and it takes time re-run the workflow. It's not rare that once the decision makers are looking at the report, there is new data available, or a more precise modelling tool, or even a different priority.

Slide 14



SpaDES helps making this workflow fluid but not disjunct, allowing for changes at different points of the workflow without compromising the whole system. This improves the speed to go from data to decisions.

Slide 15

Is SpaDES for me?

Asking and answering the right questions



15 of 52

Now the question many of you might be asking is: is SpaDES for me?

Is SpaDES for me?

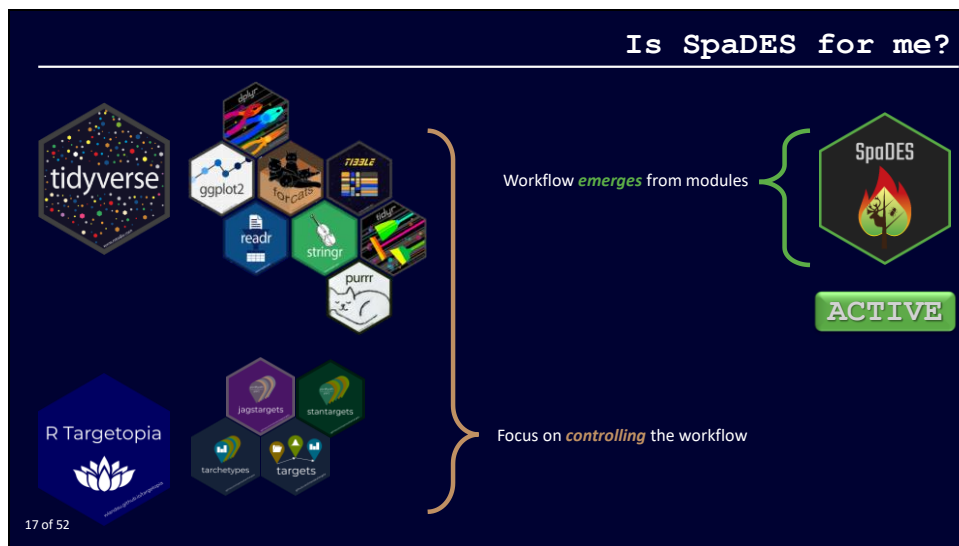
♠ **SpaDES**: much easier to learn with **R** knowledge

- ✓ Forecasting or landscape modelling are a part of my work
- ✓ My analysis scripts are very long or subdivided in many parts or steps
- ✓ I want others to be able to understand, reproduce and reuse my work
- ✓ I want to integrate my work with other researchers'
- ✓ I want to be able to reuse and build on other researcher's works
- ✓ I could use help to make my work more transparent and accessible

16 of 52

As with any new tool, there is a learning process with SpaDES. However, this process is substantially reduced if the user has a good knowledge of R. Even though SpaDES can support modules developed internally in any language (as long as the module was wrapped in a SpaDES template in R), the whole platform was built in this language, which makes projects completely transparent, accessible, and possible to debug. So I prepared a questionnaire to help you decide if SpaDES is for you. QUESTIONS. If you answer YES to one or more of these questions, SpaDES might be a good tool for your workflow.

Slide 17



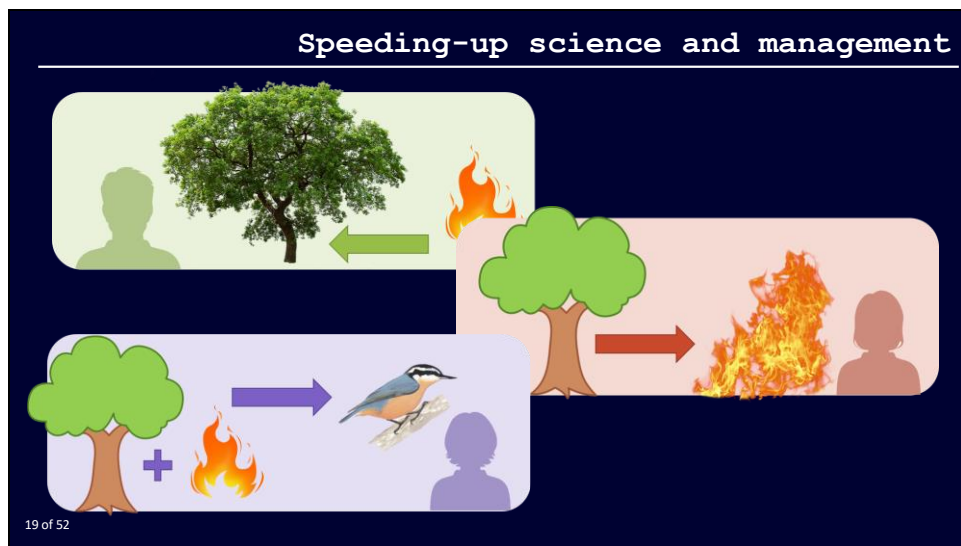
It is very important to highlight that SpaDES is not the first, nor the only tool available that focuses on helping scientists build and PERFECT workflows. Two good examples are Tidyverse and targets. The main difference between these packages and SpaDES is that, while in the first ones the **user** is responsible for **controlling** the workflow, in SpaDES the workflow **emerges** from the modules themselves created by module **developers**, not necessarily the user. This means that the user does not need to understand in detail modules that are part of their workflow to be able to use them. In other words, each module is self-explanatory to SpaDES. It says how it should run, what data it needs, what data it produces, and the workflow emerges from these data dependencies. The user can just watch what unfolds when putting modules together. Lastly, SpaDES is a long-term project and is in active development, which means it has a community ready to help.

SpaDES can speed-up science and management

Sharing your work is good for you, for science and for the world

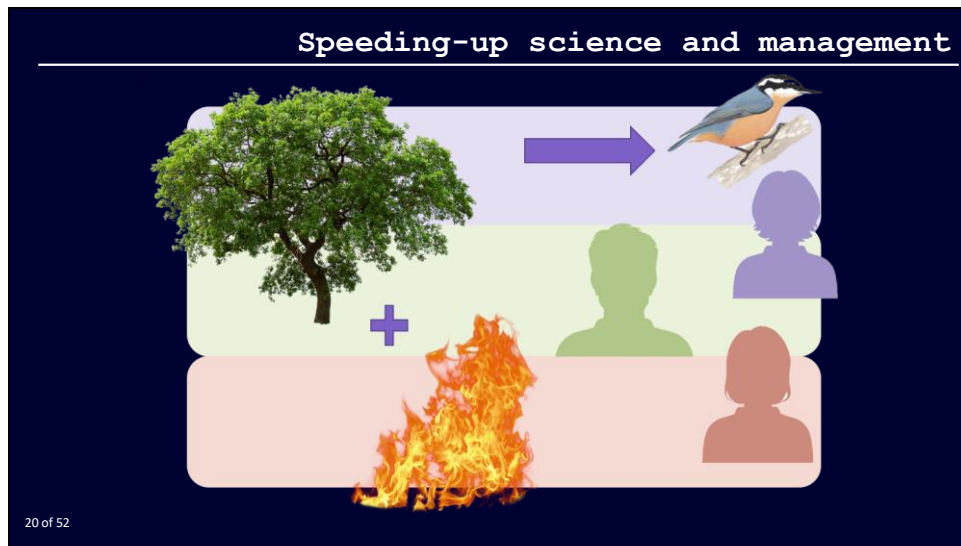


Slide 19



To understand how Spades can help speeding up science and management, it is imperative to understand Spades' inherent modularity. If you want to model forest growth and mortality on landscapes that naturally burn, and you are a forest ecologist, there is a good likelihood that your forest model will be considerably more developed, while the wildfire model might be simplified, as this is not your field of work. On the other hand, if you are a wildfire ecologist, chances are that your forest model will be simplified while your fire model might be more developed. If you are a forest bird ecologist, however, it is likely that both landscape models will be simplified while your species models will be more developed.

Slide 20



However, if all researchers are using SpaDES, the more developed models can be shared in a way that all three researchers can benefit from state-of-the-art models from outside their area of expertise.

Slide 21

Speeding-up science and management



Wildfire Models

Forest Growth and succession Models

Climate Models

Wildlife Models

21 of 52

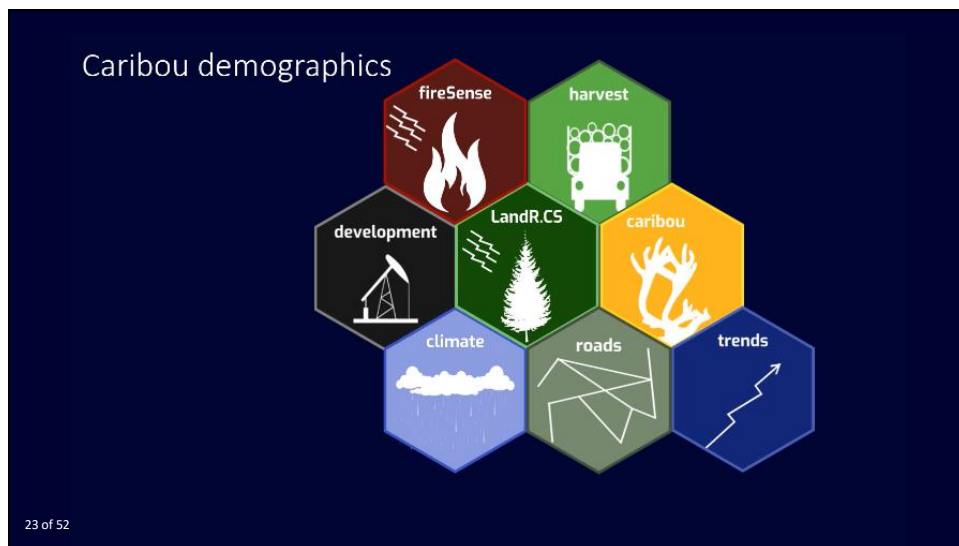
In other words, models coming from different areas of knowledge can be combined to answer important scientific and management questions.

Slide 22



For example, if we have a large collection of interoperable modules, we can quickly answer different questions by just reshuffling and recombining these pieces.

Slide 23



For example, we can answer questions about caribou demographic trends by combining landscape modules for forest growth and mortality, wildfire, different types of human disturbances, and climate change.

Slide 24



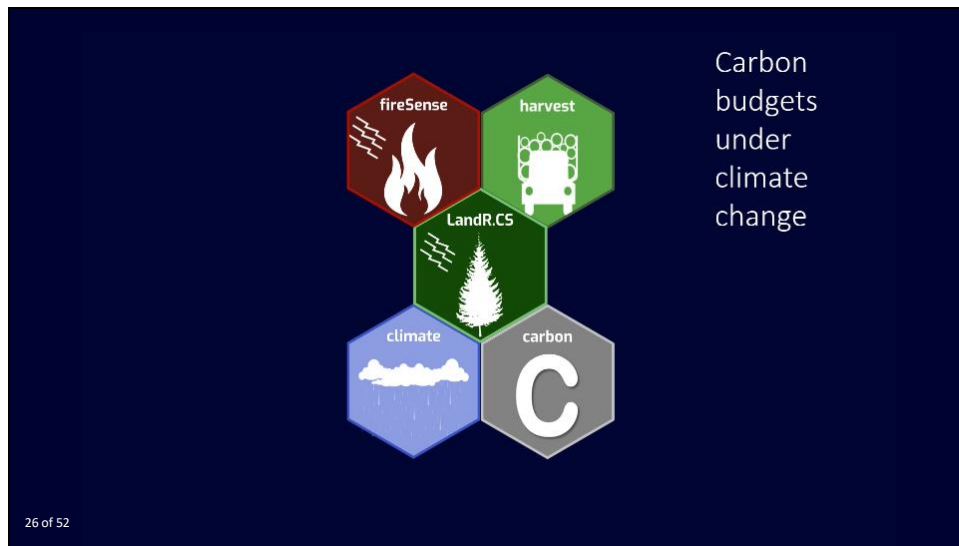
Or we can use the same landscape module collection to answer questions about caribou resource selection.

Slide 25



We could also to this, weave in Indigenous Knowledge...

Slide 26

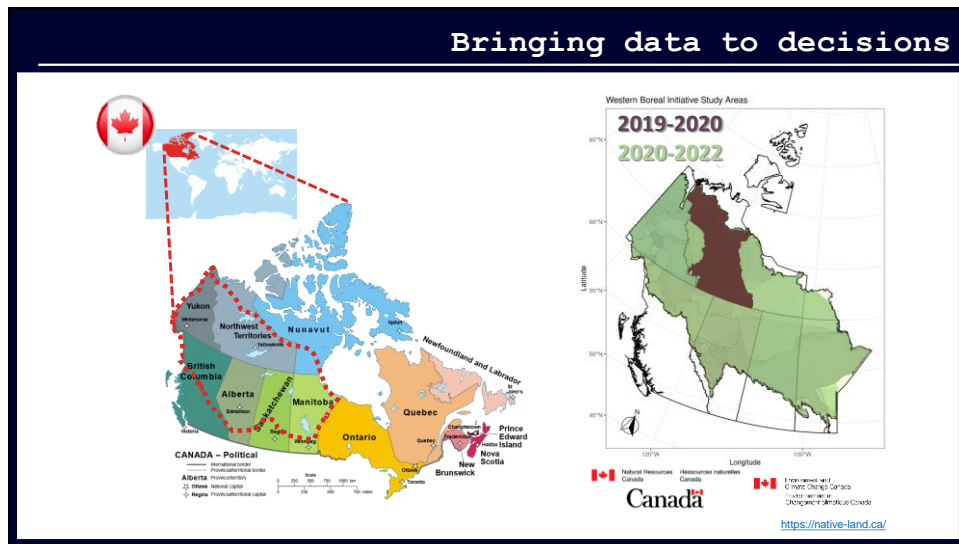


...or with the forest growth and mortality, wildfire, harvesting and climate change modules, we can answer questions about carbon budget. The possibilities to explore scientific and management questions are endless.



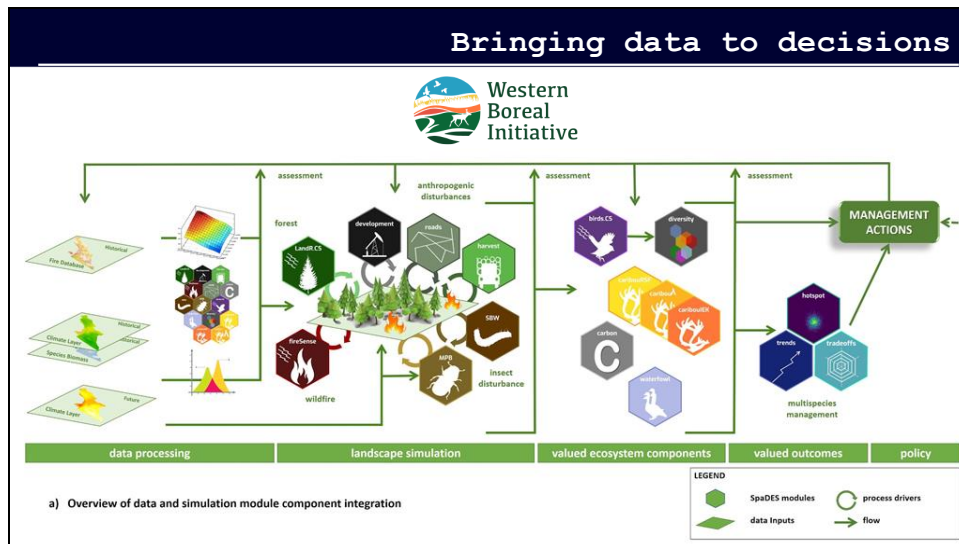
And that is what we have been doing in Northern Canada within the Western Boreal Initiative, using SpaDES to help bringing data to decisions.

Slide 28



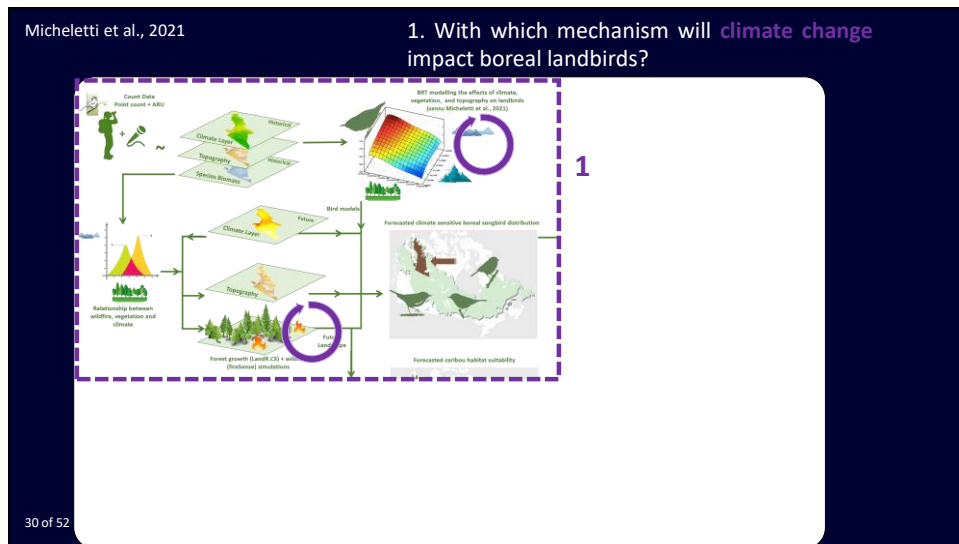
The Canadian Western Boreal Forest is located in the North-western part of the country. From 2019 to 2020, we ran a pilot project in the Northwest Territories, focusing mostly on forecasting songbirds and boreal caribou habitat and demography under changing climate. The project worked so well, that in the two subsequent years, we expanded the area almost 5 times, to cover the whole Western Boreal Forest...

Slide 29



... and we expanded from landscape, birds and caribou to harvesting, pests –spruce budworm and mountain pine beetle – carbon, and waterfowl.

Slide 30




Our framework also allowed us to ask and answer questions such as “with which mechanism will **climate change** impact boreal landbirds?”, which led to asking

Slide 31

Raymundo et al., 2024

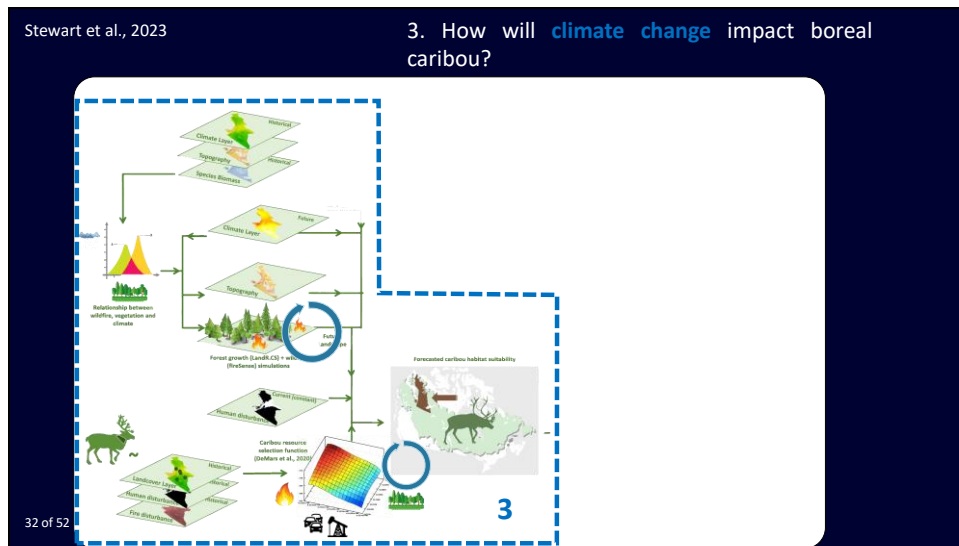
2. How will boreal landbird **abundance and distribution** change throughout the century?



31 of 52

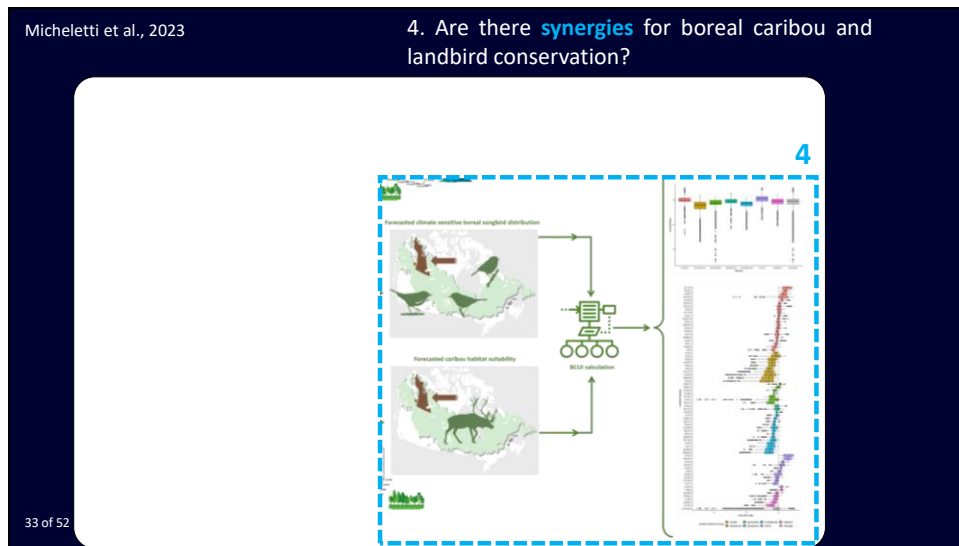
“How will boreal landbird **abundance and distribution** change throughout the century?”, and

Slide 32



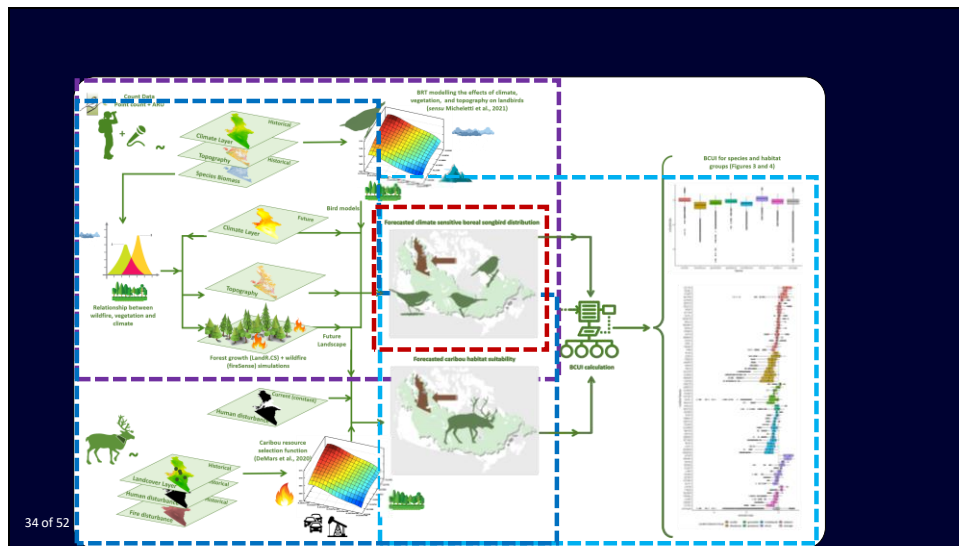
“How will **climate change** impact boreal caribou?”, to

Slide 33



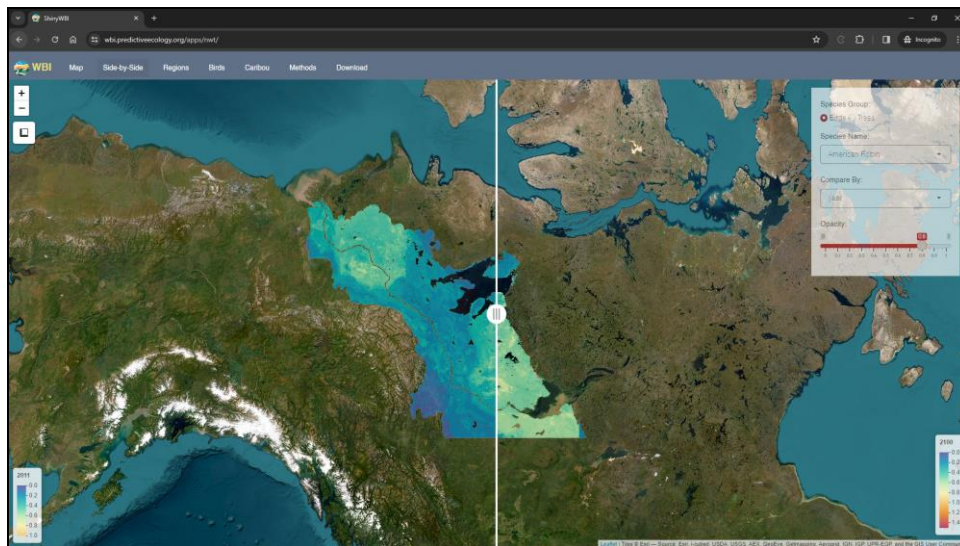
putting several of these answers together and asking “what are the **synergies** for boreal caribou and landbird conservation?”

Slide 34



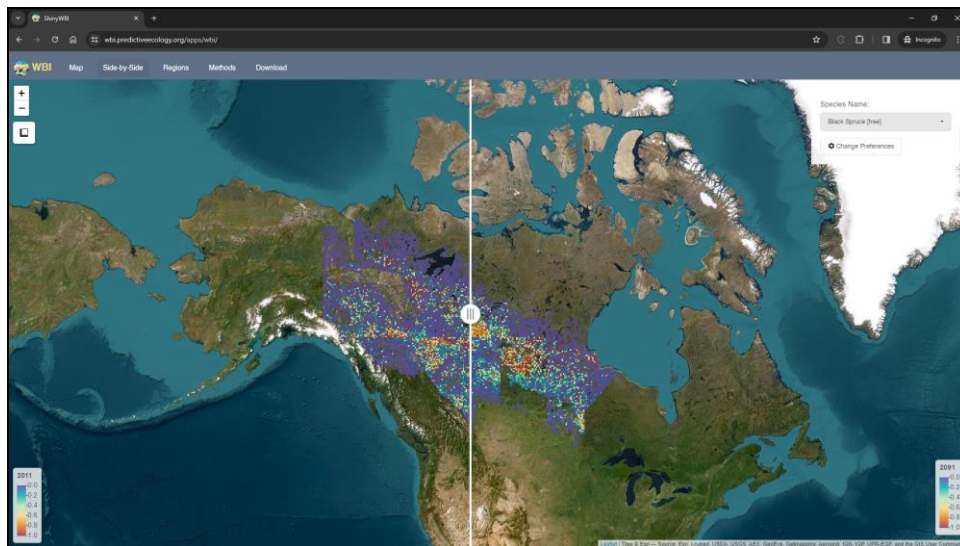
Here we see that due to its focus on modularity, SpaDES allows for the appropriate level of complexity for each project. We can have from simple projects to highly complex projects simply by adding more modules representing either previously inexistent or simplified processes.

Slide 35



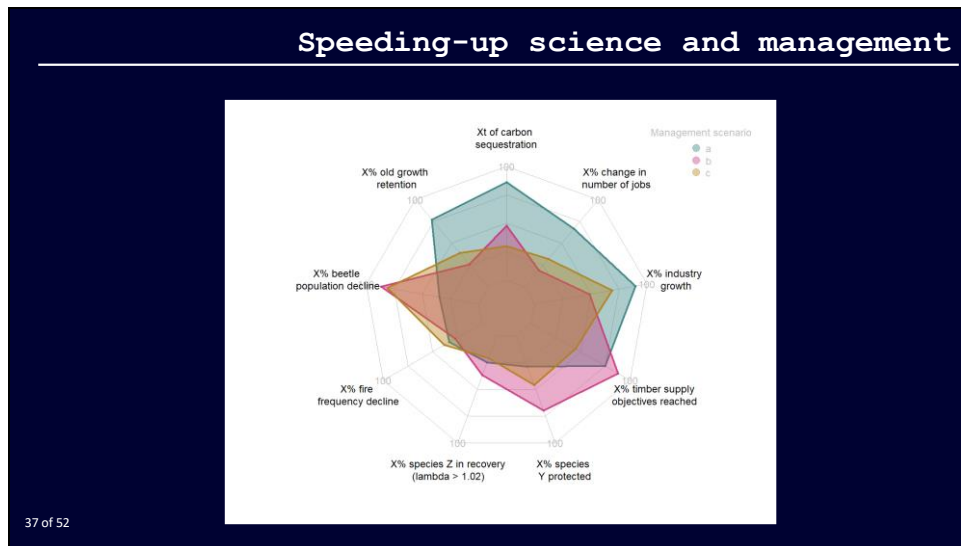
The results of these sub-projects are then presented in a Shiny app, which is helping guide management decisions for this area at both local...

Slide 36



... and regional level.

Slide 37



Our goal is to allow managers to identify the potential synergies and tradeoffs among different management scenarios and make decisions based on the best data and models available. And SpaDES can help building this framework.

SpaDES can help you
PERFICT your
workflow

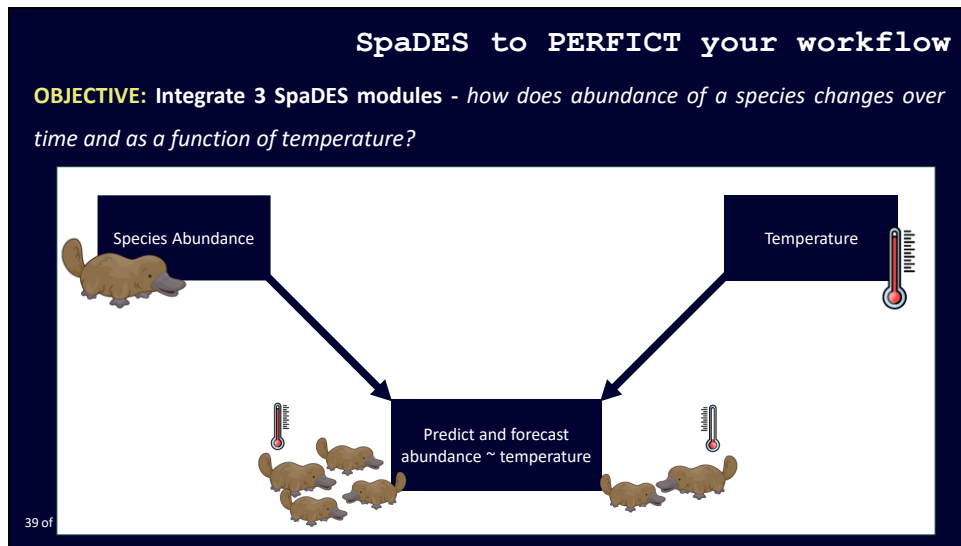
Quick hands-on demonstration of SpaDES potential



38 of 52

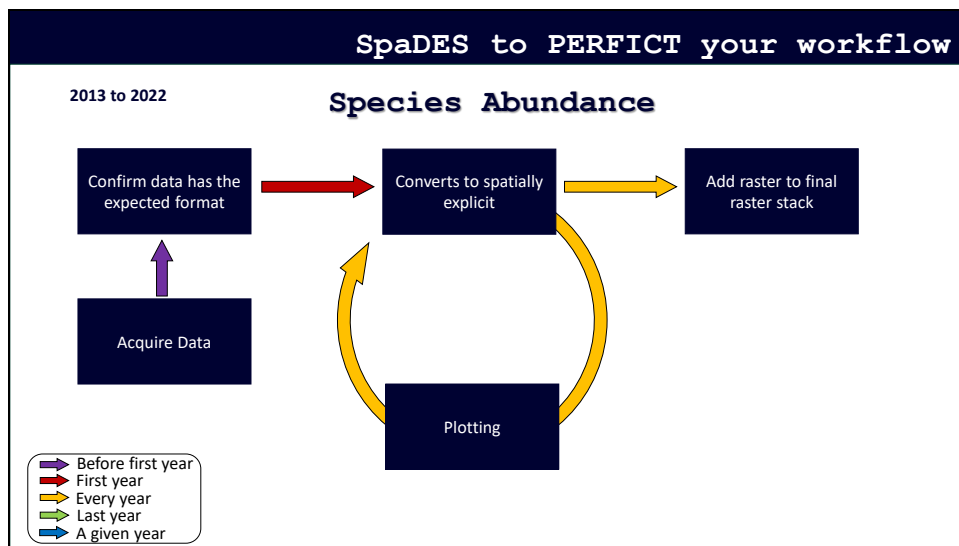
So, now that we have seen all the potential in using SpaDES, I will demonstrate a very simple example of how SpaDES can help you PERFICT your workflow.

Slide 39



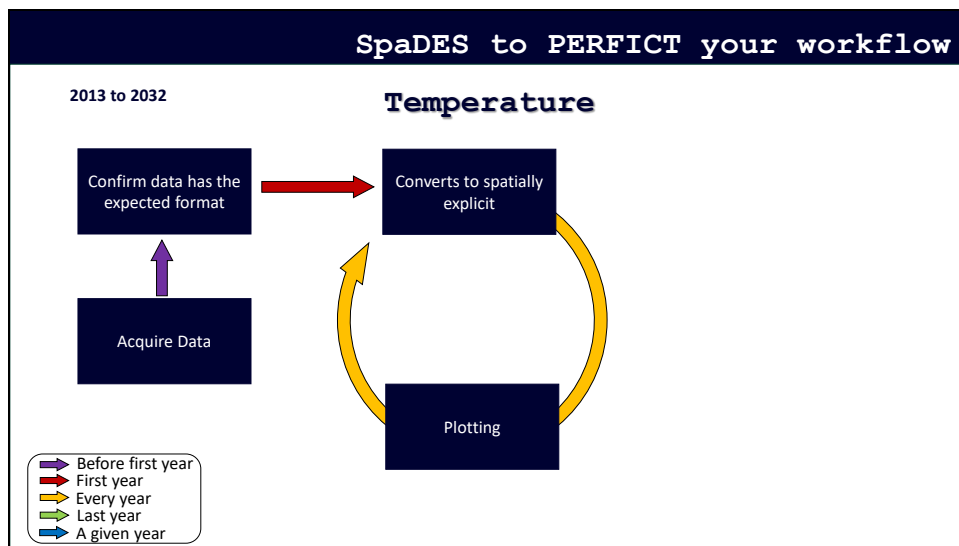
The objective of this simple project is to integrate 3 SpaDES modules and answer *how does abundance of a species changes over time and as a function of temperature?* We will then combine a Species Abundance Module of, let's say, Platypus (just because they are adorable), with a temperature module, and then fit a linear model to predict and forecast platypus abundance as a function of temperature to identify any relationships.

Slide 40



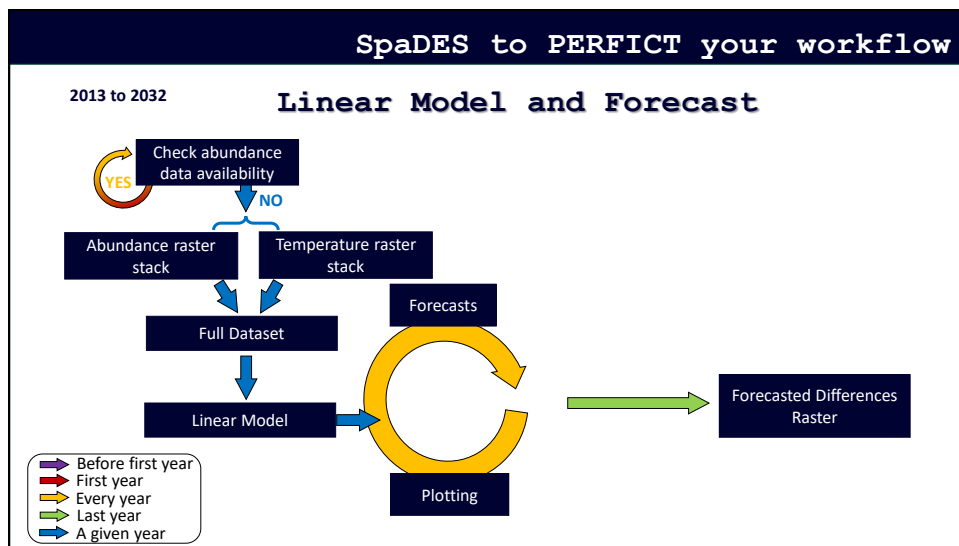
The species abundance module will run from 2013 to 2022, which are the years for which we have data. Before the simulation actually starts, the module will download data (in this case, because the user did not provide it), and then confirm that this data, which should be a data frame, has the expected format. Then in the first year, the module will convert the table of coordinates, years, and counts to a raster (which is a graphic representation of this data in the form of a grid of pixels which have the individual's counts as their values). In the same year, this raster will be appended to a list that will contain rasters for all the years (or technically speaking, a raster stack), and the module will plot the raster data. This will be done for every year data is available.

Slide 41



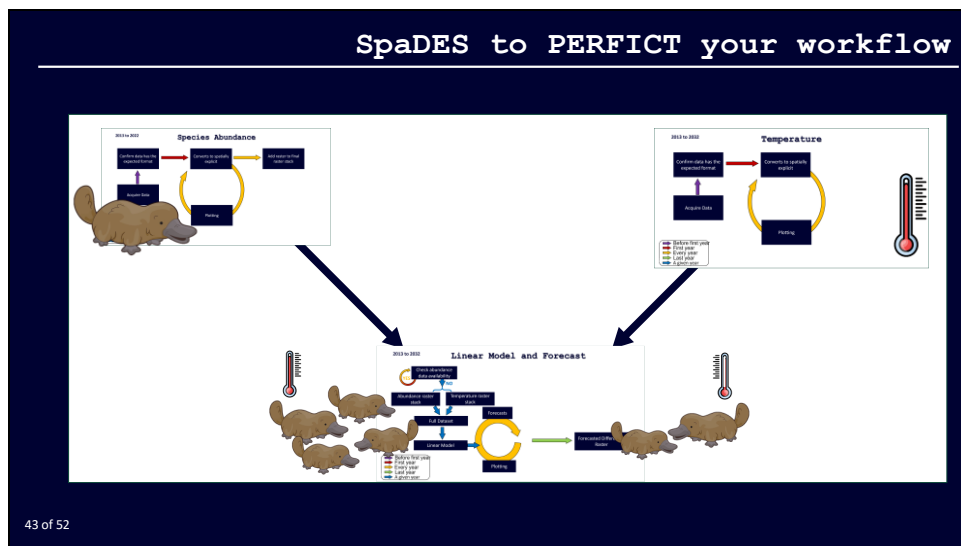
The temperature module will run from 2013 to 2032, which are the years for which we have data AND temperature forecasts. Before the simulation, the module will also download data, and then confirm that this data has the expected format. It will convert the table of coordinates, years, and temperatures to a raster, and append it to a list that will contain rasters for all the years. The module will then plot the raster created. This will also be done for every year data is available.

Slide 42



Our last module, the Linear Model and Forecast, starts in the first year by checking if data for abundance is available. The first year of which abundance data is not available, the module uses the data from both previous modules up to that point (i.e., 2022), and fits a linear model to it. Then the module performs forecasts for each year abundance data is not available (i.e., after 2022) and plots the results. This is done every year until the last one (2032), when the module calculates the differences between the last year's forecasts and the first year of abundance data, and plots it.


Slide 43



Here we can see the full workflow, now in detail.

All Hands on Deck!

Integrating simple models with SpaDES' help



44 of 52


And, now, we can finally run our code and see SpaDES in action.

All Hands on Deck!


♠ **PART I:** install the installers

♠ **PART II:** downloads modules and installs the packages required by these

♠ **PART III:** run spades



tinyurl.com/webinarEF1
demo.R
tinyurl.com/webinarEF12
exploringSpaDES.html



45 of 52

For those of you who still didn't download the material and want to try it with me, you can find the code I will run now in this URL or QR code. The code you are seeing is divided in 3 parts: The first, deals with **installing the installation packages**. The second part **sets the paths, downloads the modules and installs the packages required by each module**, which it gets from the module's metadata. The third part, is, finally, the spades call. If you haven't downloaded the files yet, please get them now. You can use these short URLs to get them. The first, demo.R can be just copied and pasted in a new Rscript. I will use Rstudio for it.

In RStudio: *Please run the file demo.R and follow with changes from the file exploringSpaDES.html (or exploringSpaDES.pdf).*

All Hands on Deck!

Predictions: completed project

Evaluated: evaluation module


Reusable: change in module's code

Freely accessible: download and full code access

Interoperable: stand-alone and module integration possible

Continuous Workflow: from data to completion in two function calls

Tested routinely: module integration tests



46 of 52

Believe it or not, with these two function calls, we demonstrated all principles from PERFICT in this simple example:

Predictions: when we saw all the completed projects with their respective forecasts

Evaluated: when we added the model evaluation module

Reusable: when we changed a module's code to generate another object that was of our interest

Freely accessible: when everyone who tried downloaded the modules and got full access to their source code. There are no secret handshakes here

Interoperable: when we saw the workflow seamlessly integrate and remove modules as needed

Continuous Workflow: when we saw the whole workflow happening from data to completion with the help of two function calls

Tested routinely: when I showed the module integration tests in the tests folder

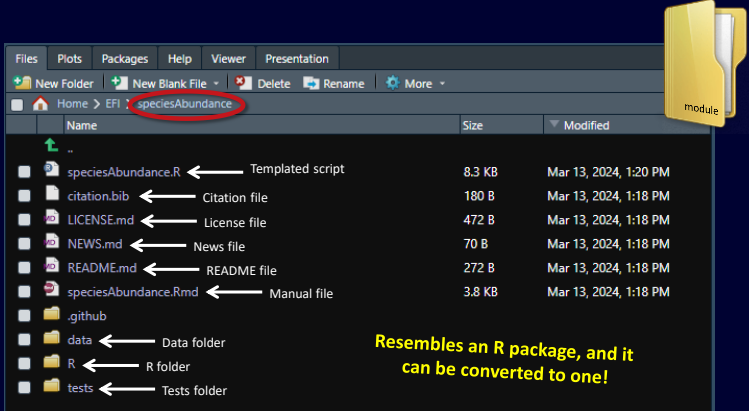


So, let's just quickly unpack a bit of SpaDES for those wanting to look under the hood.

Slide 48

SpaDES under the hood

SpaDES module structure:



Name	Size	Modified
..		
speciesAbundance.R	8.3 KB	Mar 13, 2024, 1:20 PM
citation.bib	180 B	Mar 13, 2024, 1:18 PM
LICENSE.md	472 B	Mar 13, 2024, 1:18 PM
NEWS.md	70 B	Mar 13, 2024, 1:18 PM
README.md	272 B	Mar 13, 2024, 1:18 PM
speciesAbundance.Rmd	3.8 KB	Mar 13, 2024, 1:18 PM
.github		
data		
R		
tests		


48 of 52

This is what the most important part of SpaDES, a module, looks like: for those who are acquainted with R Packages, this might look familiar. And it should. SpaDES' module's structure reminds us of R packages because R packages have a good level of FAIRness embedded in them. They contain a templated script for the module, Citation, license, news, readme, module manual, and a directory for functions called R, one for data and one for tests. Interestingly, SpaDES modules can be converted into packages, which allows them especially to be routinely tested.

SpaDES under the hood

♠ How to create a new module?

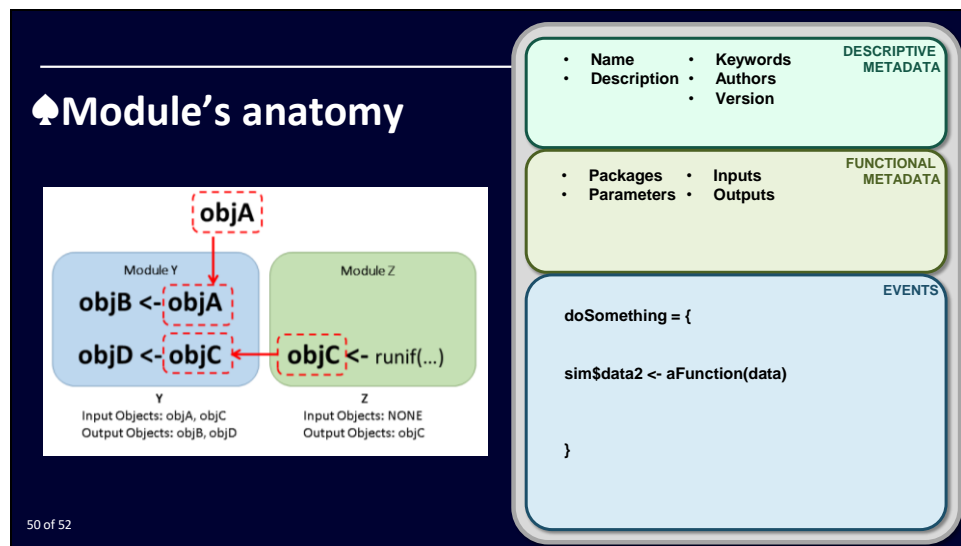
```
newModule(name = "moduleName",  
          path = getwd())
```



49 of 52

So now you might be asking How do I create a new module, then? So, pay attention: we will write **newModule()**. That's it. Let's try it out.

In Rstudio: Now you can see the instructions on all files created. If you open the myCoolNewModule.R file, you will see the template that should be filled. The module's template has two main parts. The first part is the metadata (subdivided into descriptive – human readable – and functional – which is actually hybrid, both human and machine readable) and wrapped in the function **defineModule()**. The second part is where the action happens, the events' session, defined with the function **doEvent()**.



This is the simplified anatomy of a module. It has the two parts I mentioned, the metadata, and the events. The metadata can be again subdivided in two parts. The descriptive metadata contains information such as module name, description, authors, etc., The functional metadata, is a hybrid metadata with both machine and human readable information. It contains information such as needed packages, parameters, and inputs and outputs. These are how the module communicate among themselves. For example, if we have module Y and Module Z, and module Y has inputs like Object A and C, and outputs Objects B and D, it can use Object C coming from module Z, as Object C is an output of this module. While the metadata informs SpaDES about the module's needs and outputs, the events is the section responsible to creating these outputs and, importantly, rescheduling itself or other events. One more important thing to mention is that this whole system works based on a list (a special one, but a list nonetheless). This list is called the `simList`, and it holds all the information needed for SpaDES to schedule the events and modules' orders. This means that if an event needs to use an object produced by, for example, another module, it will find this object in the `simList`. If an event is creating an object, it should store it in the `simList`. That's how modules can use and modify each other's objects – for example, a


landscape where a forest grows, and fire burns it, and then it grows back with other properties. Inside the events, you can build and use ANY functions you would on a normal R script, even if it means calling functions in python with the package reticulate, or in C++ with the package Rcpp. All you need to remember is that the order of the execution of event will be given by the order of scheduling (with possible fine tuning using a priority ordering for events happening at the same point in time).

Slide 51

Take-Home Message

- ♠ **SpaDES**: a tool that can help implement the **PERFICT** principles to modelling workflows
- ♠ In **SpaDES**, the workflow emerges from the module's metadata, regardless of how complex it is
- ♠ Module quality is responsibility of module developers, not users

SpaDES supports the development of **PERFICT FAIR** workflows with the right level of complexity and flexibility, helping researchers bring data to decisions.



Take home messages: SpaDES is a tool that can help implement the PERFICT principles to modelling workflows. In SpaDES, the workflow emerges from the module's metadata, regardless of how complex the workflow is. Module quality is responsibility of developers. This means although SpaDES greatly facilitates the implementation of the PERFICT principles, it can't ensure the module developer will create a module with the PERFICT principles. And lastly, SpaDES supports the development of PERFICT FAIR workflows with the right level of complexity and flexibility, helping researchers bring data to decisions.

Slide 52



This is what I had planned to present to you today. But before we dive into the questions, I would really appreciate if you all could take part into a short survey to help us develop and improve SpaDES tutorials and vignettes. The survey can also be done anonymously, and used to subscribe to our mailing list and discussion group. You can find the link and the QR code on the slide. Thank you very much!