Unleashing international training potential with a *ModernStats Carpentry*

Presented by Eric Anvar (OECD) and Andrew Tait (UNECE)

2022 Workshop on the Modernisation of Official Statistics

November 22, 2022

The "Meta-academy" as defined in 2022

Absence or limited…

- …common understanding of the training needs
- ..shared methodology to create learning content
- ..forum or community for trainers and academy managers to exchange practices, share on priorities
- …decision to co-invest in training efforts
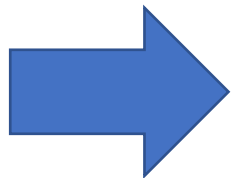
Removing those barriers defines the notion of "Meta-Academy"

Carpentries identified as a promising model

THE CARPENTRIES

A new Carpentry?

ModernStats Carpentry

Data Carpentry workshops should contain the following:

- A lesson on data organisation
- Any three other lessons from in the same domain in the Data Carpentry curriculum

Library Carpentry workshops should contain the following:

- Any three lessons from the Library Carpentry curriculum

Software Carpentry workshops should contain Software Carpentry lessons on:

- The Unix Shell
- Version Control with Git
- A programming language (R or Python)

A community of 4000 certified trainers; 90000 learners participated in 3500 workshops from #100 members (universities and libraries mostly).

How could ModernStats Carpentry help us?

Methodology to (co)create quality training

The Carpentries have a standard Lesson Program Incubation to create new training

Their model emphasizes quality assurance of lessons, through an iterative approach based on experience gained with the other Carpentries initiatives.

Forum for trainers

'Train the trainers' programme and mentoring groups results in the development of a consistently skilled network of trainers
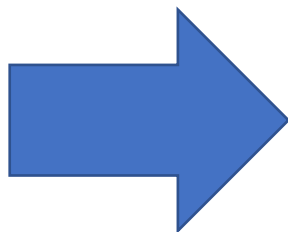
ModernStats trainers would be added to an existing network of 4000 trainers

Evolve with the industry

Through the community of trainers and maintainers that includes academia and industry partners

The Carpentries has a requirement and method for training to be kept up to date

Outcome: unlock national programs and connect them as an open ecosystem

2023 project deliverables

Work Package 1 to put together the initial ModernStats Carpentry Lesson Program *or curriculum. This work package to focus on repurposing existing Carpentries content for select key personas within statistical agencies as well as explore how to put traditional official stats course into the Carpentries framework. Under this umbrella, member NSOs encouraged to take the Carpentries' Train the Trainer course, at their own expense, to undertake pilot training sessions and help evaluate the value of the Carpentries, and determine how best to turn Carpentries content into an official stats content.*

Work package 2 to explore membership, collaboration and organizational models *between the HLG-MOS and the Carpentries. Topics to be covered include membership with the Carpentries, procurement, administrative overhead, IP of content, legal aspects, etc. The goal of this work package is essentially to determine a cost effective and sustainable business model for the HLG-MOS ModernStats Carpentry and its members to benefit from the Carpentries model and products in the context of the HLG MOS existing governance and structures (especially, the Capabilities and Communication Group).*

WP1: A quick tour in the lesson incubator…

This lesson is still being designed and assembled (Pre-Alpha version)

Home   Code of Conduct   Setup   Episodes ▾   Extras ▾   License   Improve this page ✏

Search...

Python for Official Statistics

Python for Official Statistics will teach participants the basics of Python for its use in creating Official Statistics. Participants will learn basic programming principles, and employ them in the manipulation of data and data structures.

✹ Prerequisites

FIXME

Schedule

	Setup	Download files required for the lesson
00:00	1. Introduction	What is programming? How do I document code? How do I find reliable and safe resources or code online?
00:15	2. Python Fundamentals	
00:45	3. Data Transformation	
01:45	4. List and Dictionary Methods	How can I store many values together? How can I create a list succinctly? How can I efficiently access nested data?
02:25	5. Loops and Conditional Logic	How can I do the same operations on many different values? How can my programs do different things based on data values?
03:25	6. Alternatives to Loops	How can I vectorize my loops?
03:55	7. Creating Functions	What are functions, and how can I use them in Python?

Visit: https://unece.github.io/ModernStats_Python/

WP1: History of the work package

Initial Aims

- Develop 3 lessons with several episodes each

  ➢ The lessons would focus on **Python**, **R**, and **Git**

- Bring the lessons to alpha stage by summer

Issues

- Departure of previous WP1 lead Kate Burnett-Isaacs

- Some loss of contributors due to workload increases, leave, or change of jobs

- Change in the official Carpentries content format (from "styles" to "workbench")

WP1: Current status of the work package

Replacement of Kate Burnett-Isaccs

- Jonathan Wylie of StatCan has agreed to take over from Kate

Shift in Goals and Timelines

- Due to the Python lesson being the most complete, efforts are now focused on that lesson

- The current primary aim is to bring the Python from pre-alpha to the alpha stage

- Restarting development on the other lessons is under discussion

- The Carpentries Org is creating a simplified guide on transitioning to the new format

WP2: Five questions

Question 1: How can we incentivize trainers to fully embrace the framework?

Question 2: Is the Carpentries IP model acceptable for public sector statistical organizations?

Question 3: How does the governance of the Data and Software Carpentries work?

Question 4: Are we comfortable becoming direct members of the Carpentries?

Question 5: Do we all agree on the preferred, most realistic approach?

Question 1: How can we incentivize trainers to fully embrace the framework?

There is a cost adopting a common framework vs. doing it your own way but there are advantages...

Carpentries lesson infrastructure, lesson program model, instructors' training

- Leverage proven tools and teaching methods

- Emphasis on effective teaching and helping with curriculum design

- Reuse lesson content... and be reused

- Join a global community of peer-trainers

Challenge: most experts in statistical organisations are not trained trainers/coaches. Sharing expertise is not their day job. Fully recognising the value of them sharing their expertise and supporting them doing so in consistent manner – are not necessarily high on the agenda.

Question 2: Is the Carpentries IP model acceptable for public sector statistical organizations?

Carpentries have a classical open model – which requires a transfer of IP for "official lessons"

On Intellectual Property: Content of official lessons belongs to The Carpentries, which take ownership of their maintenance as part of the official lesson program. Community-developed lessons (Carpentries Incubator) belong to their authors

On Licensing: Lesson content "prose": CC-BY; lesson content example code: MIT; example data is usually CC0. Data should be available to everyone who can access the lesson material.

Question 3: How does the governance of the Data and Software Carpentries work?

Option 1: create lessons under the existing Data Carpentries

Option 2: create a separate ModernStats Carpentries

ModernStats Carpentry
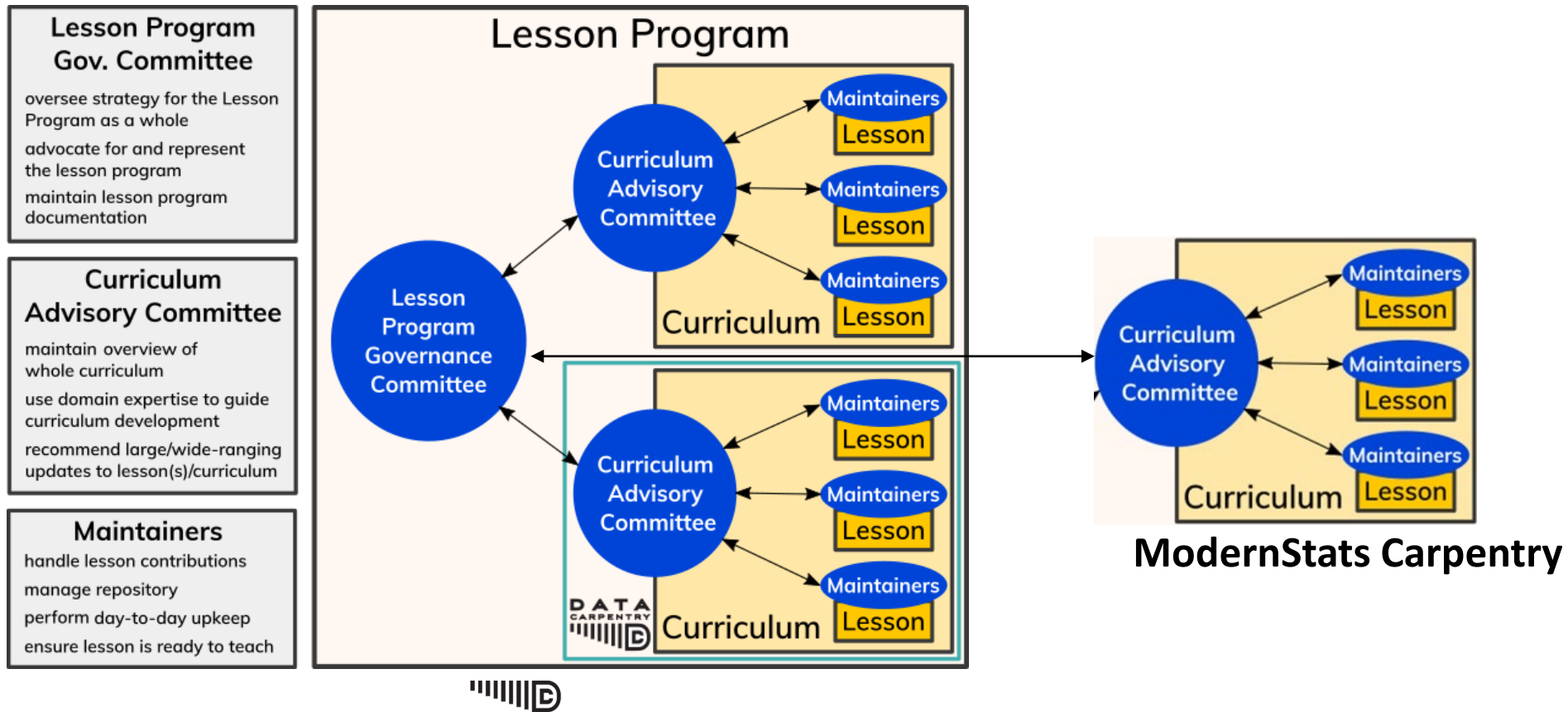
Option 1: create lessons under

DATA CARPENTRY

BUILDING COMMUNITIES TEACHING UNIVERSAL DATA LITERACY

"Data Carpentry develops and teaches workshops on the fundamental data skills needed to conduct research. Our mission is to provide researchers high-quality, domain-specific training covering the full lifecycle of data-driven research."

- 6 curricula are currently being developed: Astronomy, Ecology, Genomics, Geospatial, Image Processing, Social Sciences.
- ModernStats could contribute to under Social Science, or under a specific curriculum (for example, "official statistics").

- Process for adoption under the DC Committee
  - Raise this with the DC Governance Committee early
  - Develop material in the Incubator (eg the UNECE space used for the Python course)
  - Test your material and also gather feedback from other Instructors
  - DC Governance Committee approval step TBD
  - Establish Maintainer teams and Curriculum Advisory Committee

Option 2: create the

Possible mission statement: "ModerStats Carpentry develops and teaches workshops on the fundamental data skills needed to collect, process, analyse and disseminate official statistics. Our mission is to provide official statisticians high-quality, domain-specific training covering the full lifecycle of official statistics, on par with market trends in the area of data technologies. Promotion of data standards, especially those developed and maintained under ModernStats, constitute a specific mission to promote sharing and enforcement of good statistical standards."
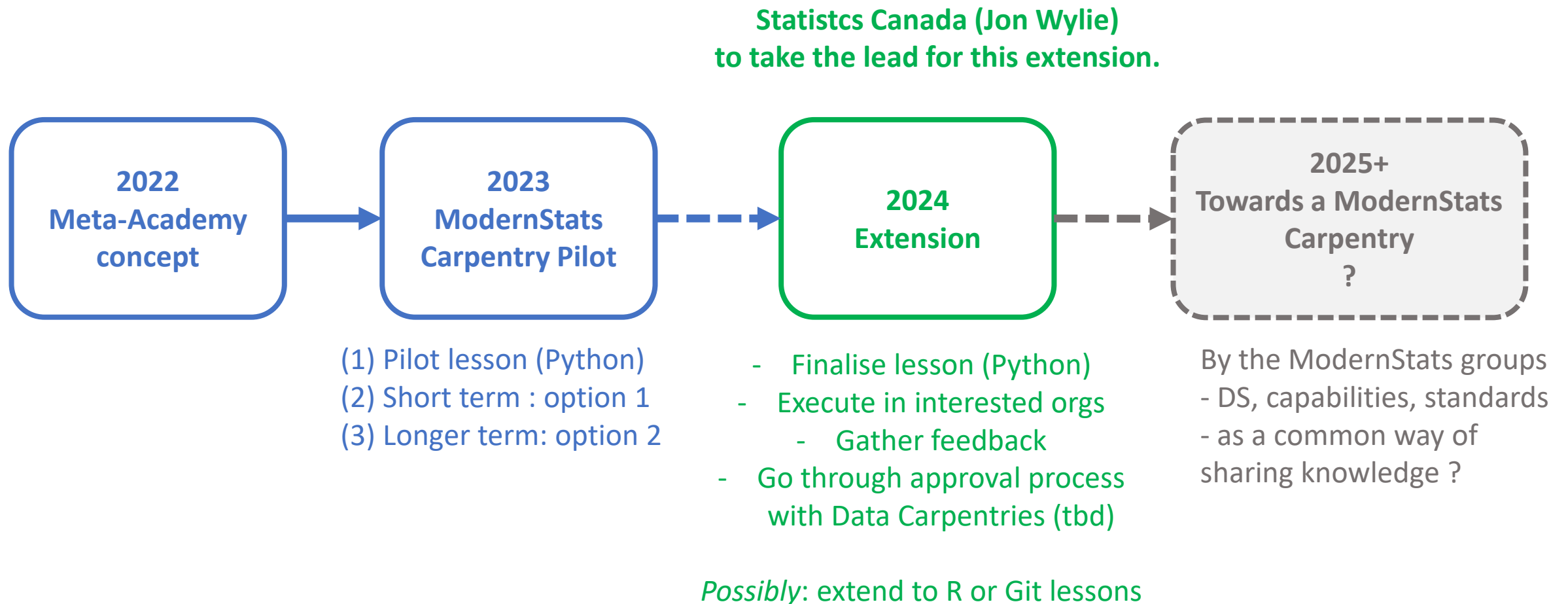
- Focus for now has been on "technical skills" eg R/Python programming or Git – which do overlap with the Data Carpentries.
- But extending the approach to domain specific skills (eg SNA), standards (eg GSBPM) and innovation developed in HLG groups (eg ML for official statistics) will likely exceed the capacity of a curriculum under the Data Carpentries.

- A much heavier process, which will take years of incubation
  - Develop material in the Incubator
  - Test your material and also gather feedback from other Instructors
  - Establish governance structures for your project – probably connected to HLG-MOS capabilities group
  - Submit application to Executive Council for Lesson Program Incubation
  - Refine curriculum and community, governance structures
  - Integrate with The Carpentries as a new lesson program
  - One precedent: the High Performance Computing Carpentry - HPC Carpentry (hpc-carpentry.org)

Question 4: Are we comfortable becoming direct members of the Carpentries?

Membership Benefits					
	Bronze	Silver	Gold	Platinum	Titanium
Number of Centrally-Organised workshops	2	2	3	0	0
Number of Instructors trained	0	5	12	15	0
Membership Fee (USD Annual) per World Bank's income categorisation					
High income	$3,600	$8,100	$16,200	$13,500	$5,000
Upper-middle income	$2,700	$6,075	$12,150	$10,125	$3,750
Lower-middle income	$1,800	$4,050	$8,100	$6,750	$2,500
Low income	$900	$2,025	$4,050	$3,375	$1,250

See Become a Member Organisation (carpentries.org)

Question 5: Do we all agree on the preferred, most realistic approach?

**Statistcs Canada (Jon Wylie)
to take the lead for this extension.**

**2022
Meta-Academy
concept**

**2023
ModernStats
Carpentry Pilot**

**2024
Extension**

**2025+
Towards a ModernStats
Carpentry
?**

(1) Pilot lesson (Python)
(2) Short term : option 1
(3) Longer term: option 2

- Finalise lesson (Python)
- Execute in interested orgs
- Gather feedback
- Go through approval process
  with Data Carpentries (tbd)

Possibly: extend to R or Git lessons

By the ModernStats groups
- DS, capabilities, standards
- as a common way of
sharing knowledge ?

Questions?

Thank you