Formal and Informal Collaboration

A presentation at Davidad's second ARIA workshop

Owen Lynch

Topos Institute

March 4, 2024

Overview

Follow along at forest.localcharts.org/aria-0001.xml!

1. Motivation
2. Informal collaboration
3. Formal collaboration

Motivation: What does success mean?

Successful implementation of Davidad's program thesis over the next 3 years implies something like

- Several new fields worth of novel math research
- $>$1,000,000 lines of code

Motivation: What does success mean?

Successful implementation of Davidad's program thesis over the next 3 years implies something like

- Several new fields worth of novel math research
- $>$1,000,000 lines of code

This is only possible if we either:

- Clone Urs and ekmett a couple of times and form them into an (aligned) borg-like mindmass, or

Motivation: What does success mean?

Successful implementation of Davidad's program thesis over the next 3 years implies something like

- Several new fields worth of novel math research
- $>$1,000,000 lines of code

This is only possible if we either:

- Clone Urs and ekmett a couple of times and form them into an (aligned) borg-like mindmass, or
- Get really good at collaboration.

Motivation: Current de-facto standards for collaboration

- Informal technical writing: overleaf+arXiv
- Technical computing: github repositories containing arbitrary code

Informal collaboration: A dream

- Monday morning (UK time): DJM writes down new definition
- Monday afternoon (EU time): Matteo adds some key lemmas
- Monday afternoon (Pacific time): Sophie spots hole
- Tuesday morning (EU time): A long-time lurker comments for the first time on idea for patching hole
- ...

Informal collaboration: A dream

- Monday morning (UK time): DJM writes down new definition
- Monday afternoon (EU time): Matteo adds some key lemmas
- Monday afternoon (Pacific time): Sophie spots hole
- Tuesday morning (EU time): A long-time lurker comments for the first time on idea for patching hole
- ...
- By Thursday night: Enough material for a paper
- Friday morning (UK time): Organize material into a paper by writing an abstract, collating background+new material into a reasonable order, and then exporting to arXiv-compatible LaTeX. All authors of transcluded material notified and given a chance to review.
- Friday afternoon (UK time): pub.

Informal collaboration: How does forester work?

Forester takes a collection of files with TeX-like syntax and produces both a static website and LaTeX.

Key features:

- Transclusion
- Linking, backlinking, and citation
- Macros
- TikZ$\rightarrow$SVG
- Customizable LaTeX export
- Better error messages than LaTeX

Informal collaboration: Just add (more) users

LocalCharts is live!

- Medium-sized forum
- Small but growing forester instance
- This talk built via forester
- Compatible with UK law for government projects

Formal collaboration: Math on the computer

What does it mean to do math on the computer?

- Logician: propositions as types.
  - Characteristic Algorithms: Martin-Lof type checking
  - Programming languages: Isabelle, Lean, Coq, Agda
- Algebraist: Symbolic rewriting
  - Characteristic Algorithms: Groebner bases, e-graphs
  - Programming languages: Mathematica, Macaulay2, OBJ3, Z3
- Engineer/statistician: Numerical computing
  - Characteristic Algorithms: Euler's method, MCMC, gradient descent
  - Programming languages: Fortran, MATLAB, Julia

Formal Collaboration: Polyglot Scientific Models

Dilemna:

1. Don't want to rewrite tensorflow
2. Don't want to write everything in Python

Solution: Models should be language-independent "initial algebras" of systems doctrines.

Algorithms are "model semantics" that apply over large classes of models.

Formal Collaboration: Models as Data

What can be cross-language?

- Algebraic Data Types
- Generic types
- Multidimensional Arrays
- Presentations of algebraic structures (i.e. ring presentations by generator and relations).
- Knowledge bases, i.e. collections of “facts” in the style of prolog.

What can’t?

- Arbitrary functions
- Arbitrary dependent types

Formal Collaboration: Implementation of "Models as Data"

1. Type theory for data
2. Embed into existing languages
3. Build storage system

Formal Collaboration: Version the Source of Truth, Cache Everything Else

- Structured version control for models?
- Version control the source of truth, which could be
  - The model itself
  - Stochastic model search + random seed
  - Textual DSL
  - A composition diagram with other models inserted
- Everything downstream of source of truth: deterministically cache
- Nix is current state-of-the-art.

Informal+Formal Collaboration: literate programming?

- I thought literate programming is dead... but is it?
  - Rustdoc is literate programming
  - 1lab is literate programming
  - PBRT is literate programming
  - Jupyter is literate programming
- Caching enables principled notebook computing
- Explainable AI involves AI... and explanations!

Conclusion

- Success requires scaling our development via more effective collaboration
- Informal collaboration needs to scale beyond “couple of mathematicians write standalone paper”
- Formal collaboration needs to scale beyond “software package for single type of model in a single language”
- Next steps: intertypes

We shape technology for public benefit by advancing sciences of connection and integration.

– Topos Institute