Computational reproducibility in theory and practice

Konrad Hinsen

Centre de Biophysique Moléculaire, Orléans, France
and
Synchrotron SOLEIL, Saint Aubin, France

19 April 2021

Why reproducibility matters

Royal Society of London for Improving Natural Knowledge (1663)

The reproducibility crisis

Not logged in Talk

Article Talk | Read Edit View history

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Current events
Random article
About Wikipedia
Contact us
Donate

Contribute
Help
Learn to edit
Community portal
Recent changes
Upload file

Tools
What links here
Related changes
Special pages
Permanent link
Page information

Replication crisis

From Wikipedia, the free encyclopedia

The **replication crisis** (or **replicability crisis** or **reproducibility crisis**) is, as of 2020, an ongoing methodological crisis in which it has been found that many scientific studies are difficult or impossible to replicate or reproduce. The replication crisis affects the social sciences and medicine most severely.[2][3] The crisis has long-standing roots; the phrase was coined in the early 2010s[4] as part of a growing awareness of the problem. The replication crisis represents an important body of research in the field of metascience.[5]

Because the reproducibility of experimental results is an essential part of the scientific method,[6] the inability to replicate the studies of others has potentially grave consequences for many fields of science in which significant theories are grounded on unreproducible experimental work. The replication crisis has been particularly widely discussed in the field of psychology and in medicine, where a number of efforts have been made to re-investigate classic results, to determine both the reliability of the results and, if found to be unreliable, the reasons for the failure of replication.[7][8]

Contents [hide]

1 Scope
1.1 Overall
1.2 In psychology
1.2.1 Psychology replication rates
1.2.2 A disciplinary social dilemma
1.2.3 "Methodological terrorism" controversy

Biophysics: bad protein structures

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction

Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools

What links here
Related changes

Geoffrey Chang

From Wikipedia, the free encyclopedia

Geoffrey Chang is a professor at the University of California, San Diego's Skaggs School of Pharmacy and Pharmaceutical Sciences and Department of Pharmacology, School of Medicine. His laboratory focuses on the structural biology of integral membrane proteins, particularly exploring X-ray crystallography techniques for solving the tertiary structures of membrane proteins that are notoriously resistant to crystallization. The laboratory has specialized in structures of multidrug resistance transporter proteins in bacteria. In 2001, while a faculty member of The Scripps Research Institute, Chang was awarded a Beckman Young Investigators Award,[1] designed to support researchers early in their academic careers, for his work on the structural biology of multidrug resistance.[2] Chang announced a move from Scripps to neighboring UC San Diego in 2012.[3]

In 2007, Chang and coauthors retracted five previously published papers describing the structures of three multidrug transporter proteins after another research group published a widely differing structure, which led to the discovery of a critical bug in the Chang group's custom software tools.[4] Since that time, however, Chang has published other papers in the field of structural biology,[5][6] and has been awarded a EUREKA grant, "for exceptionally innovative research projects that could have an extraordinarily significant impact on many areas of science," from the National Institutes of Health.[7]

We still don’t know the full truth

COMMENTARY

Five retracted structure reports: Inverted or incorrect?

BRIAN W. MATTHEWS

Institute of Molecular Biology, Howard Hughes Medical Institute, and Department of Physics, University of Oregon, Eugene, Oregon 97403, USA

B.W. Matthews, Protein Science, 2007

Chemical physics: how many phases for supercooled water?

PHYSICS TODAY

HOME BROWSE INFO RESOURCES JOBS

DOI:10.1063/PT.6.1.20180822a

22 Aug 2018 in Research & Technology

The war over supercooled water

How a hidden ~~coding error~~ methodological choice fueled a seven-year dispute between two of condensed matter's top theorists.

Ashley G. Smart

A.G. Smart, Physics Today, 2018

The trinity of reproducibility

Experimental reproducibility:
Re-do an experiment and get sufficiently close results.

Statistical reproducibility:
Re-do a study with a different sample and get sufficiently close results.

Computational reproducibility:
Re-do a computation (data analysis, simulation, ...) and get identical results.

Causes of the reproducibility crisis

Experimental reproducibility: well understood
Re-do an experiment and get sufficiently close results.

Statistical reproducibility: incomplete understanding of statistical methods
Re-do a study with a different sample and get sufficiently close results.

Computational reproducibility: lack of rigor in documenting computations
Re-do a computation (data analysis, simulation, ...) and get identical results.

Reproducibility vs. replicability

Reproducibility

- Technical
- Done right?
- Straightforward.
- Simple answer: yes/no
- **Verification**

Replicability

- Scientific
- The right thing to do?
- Difficult to judge.
- Complex answer: if... maybe...
- **Validation**

In scientific computing:

- Same software
- Same parameters
- Same data
- **Identical result?**

- New software
- Same parameters
- Same data (or not)
- **Equivalent result?**

Reproducibility vs. replicability

Scientific method: **replicability** / **validation**

Reproducibility is a **verification** criterion that accelerates validation

Why reproducibility matters in scientific computing

- Know for sure what has been computed.
- Ability to verify the computation.
- Safe reuse for other applications.

Particularity of computation: determinism

Except for parallel computing, a deal with the devil: trade speed against loss of control.

Robustness

Scientific methods should not be sensitive to technical details.

Such as: compiler versions, operating systems, numerical discretization schemes, ...

Robustness checks test replicability, not reproducibility.

What's a computation?

230 A. M. TURING [Nov. 12,

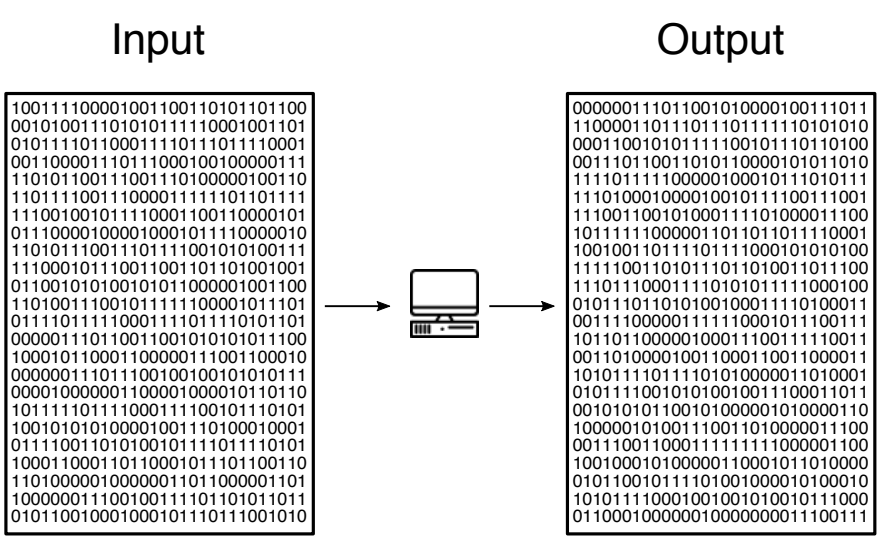
ON COMPUTABLE NUMBERS, WITH AN APPLICATION TO THE ENTSCHEIDUNGSPROBLEM

By A. M. TURING.

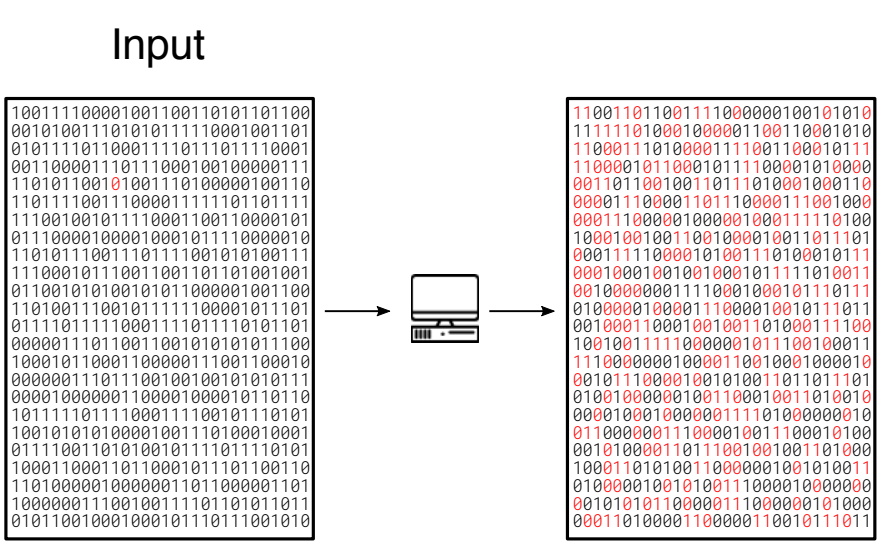
[Received 28 May, 1936.—Read 12 November, 1936.]

Proceedings of the London Mathematical Society **42**, 230 (1937)

What's a computation?

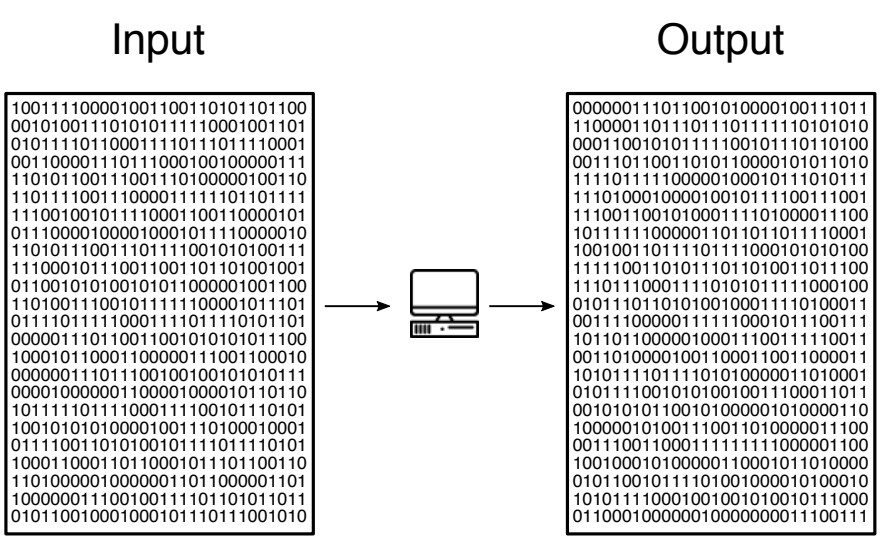


Chaos!

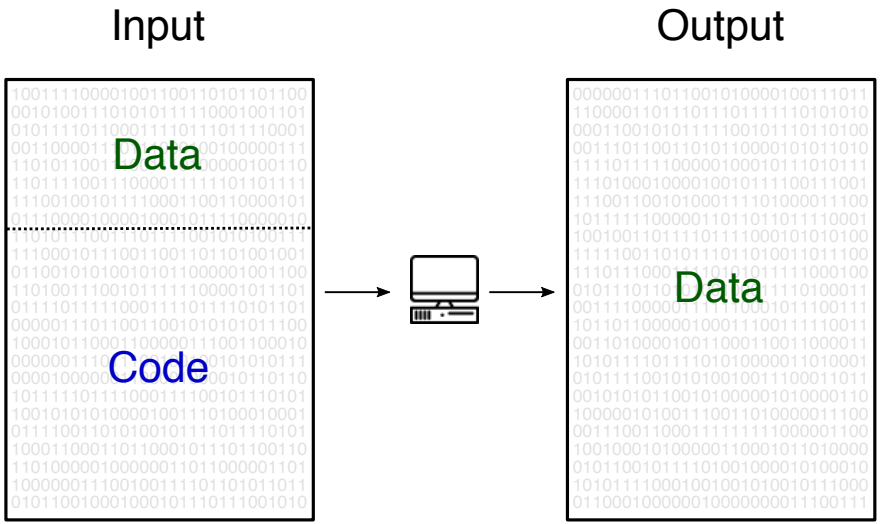


Computer by Creative Stall from the Noun Project

What's a computation?

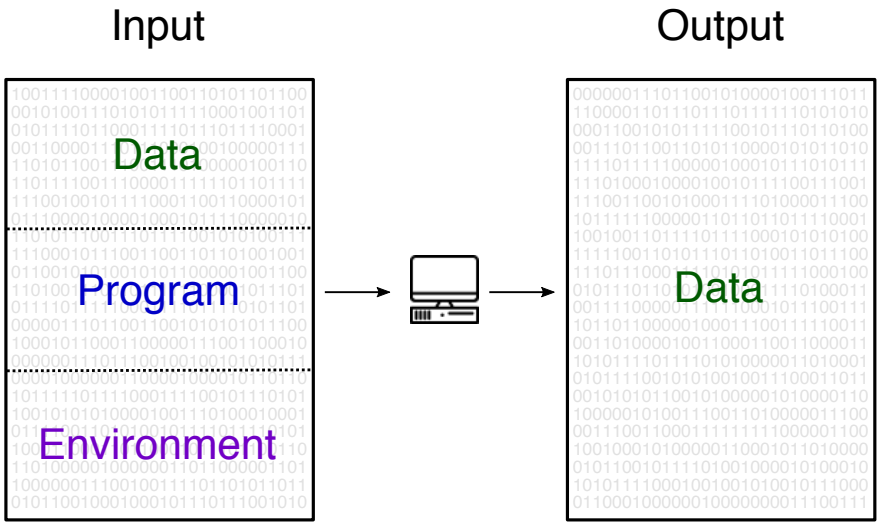


What's a computation?



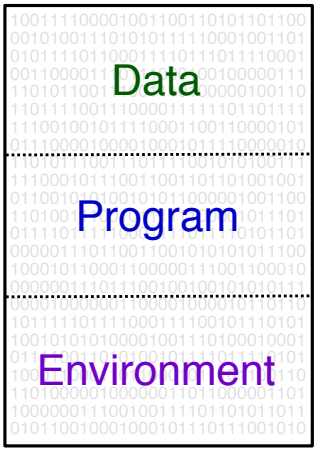
Computer by Creative Stall from the Noun Project

What's a computation?



Computer by Creative Stall from the Noun Project

The scientist’s point of view



my research

my colleagues' code

stuff I don't care about

What's the result of this program?

data_analysis.py

```
from datalib import Dataset

points = [(1, 1), (-1, 1), (2, 4)]

data = Dataset()
for x, y in points:
    if x > 0:
        data.add_value(y)
print(data.average())
```

Know your libraries *and languages*

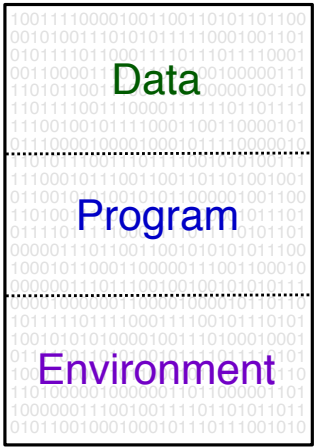
datalib.py

```
class Dataset(object):

    def __init__(self):
        self.values = []

    def add_value(self, value):
        self.values = [value]

    def average(self):
        return sum(self.values, 0)/len(self.values)
```

Surprise! A bug! `add_value` stores only the last value!

The result of `data_analysis.py` is thus 4. More precisely: 4 in Python 2 but 4.0 in

The meaning of bits

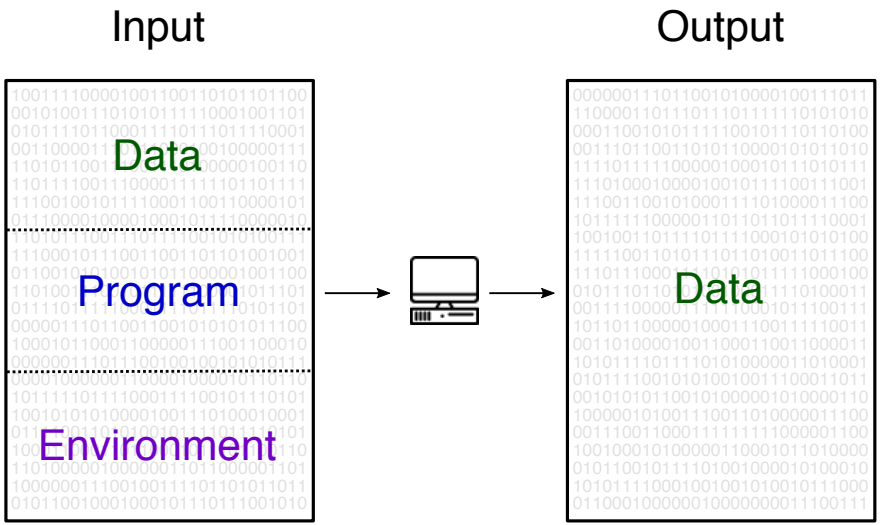


zeros and ones

interpretation of the data

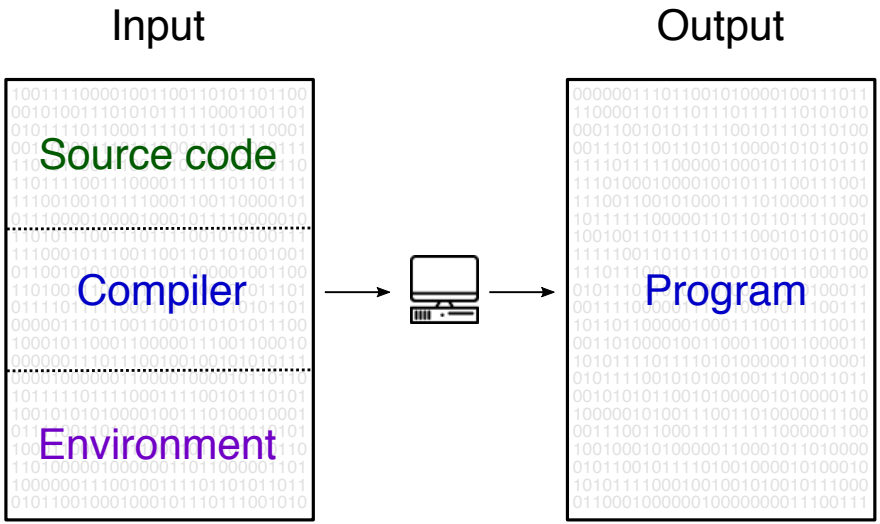
interpretation of the program

What's a computation?



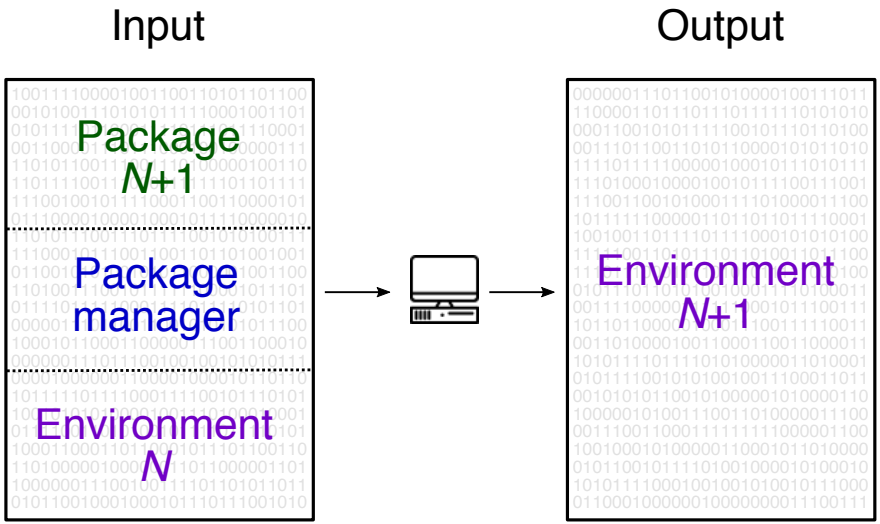
Computer by Creative Stall from the Noun Project

Where does the program come from?



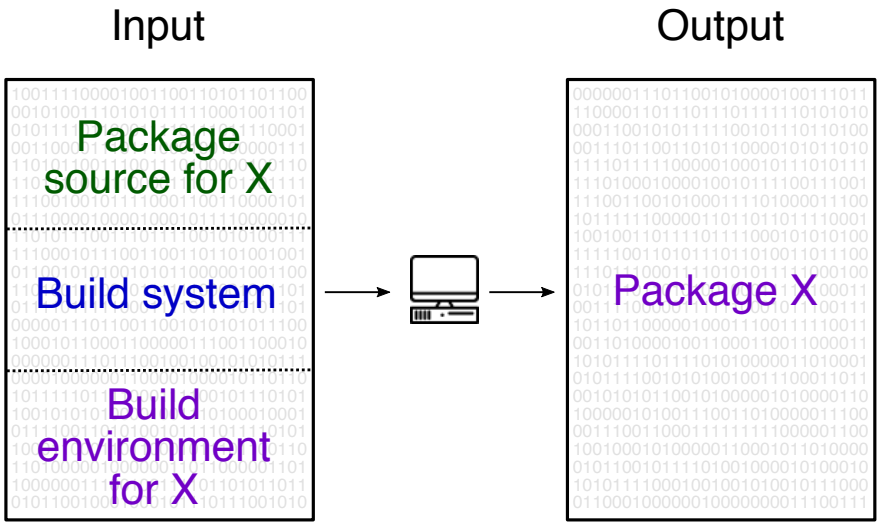
Computer by Creative Stall from the Noun Project

And where does the environment come from?



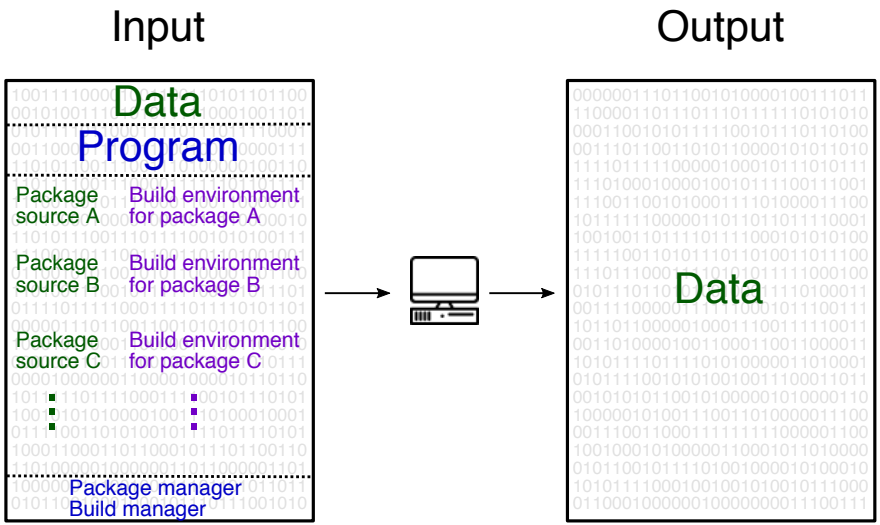
Computer by Creative Stall from the Noun Project

OK, then, where do the packages come from?



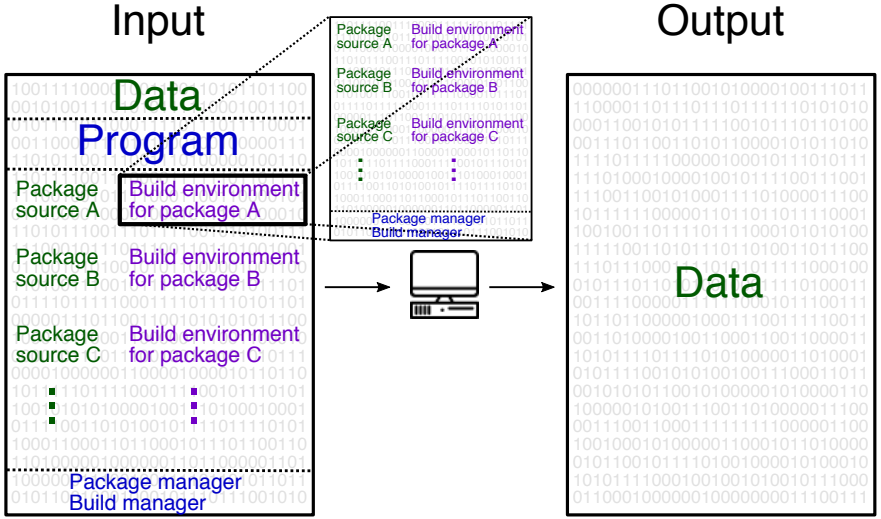
Computer by Creative Stall from the Noun Project

In summary...



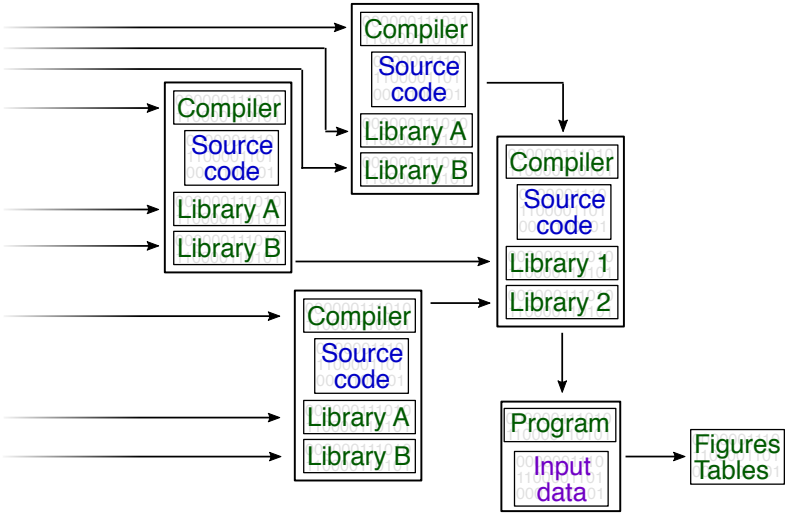
Computer by Creative Stall from the Noun Project

Without forgetting...



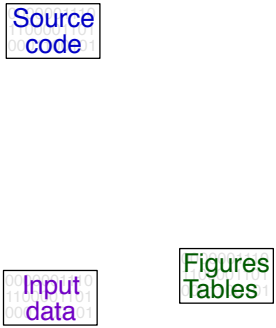
Computer by Creative Stall from the Noun Project

A different view

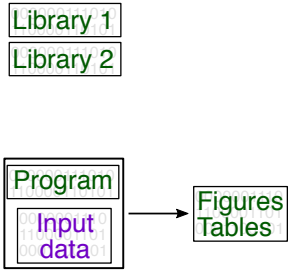


Traditional paper

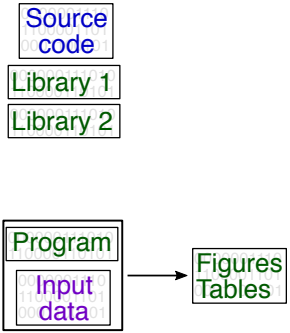
Paper with code and data



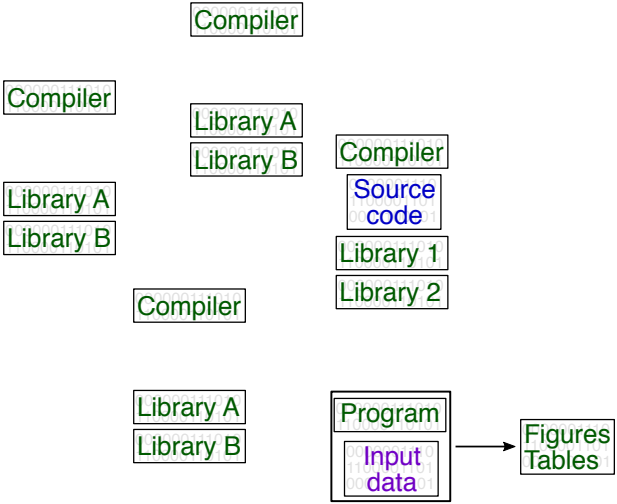
Paper plus Docker container and data



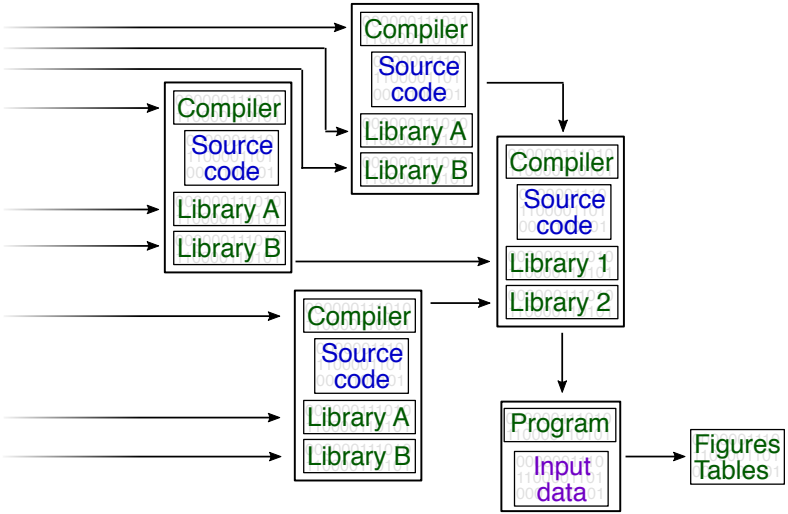
Paper plus source, Docker container, and data



Paper plus source, software packages, and data



Paper plus source, software via Guix or Nix



Case study: floating-point arithmetic

- Standardized in 1985 by the norm IEEE 754
- Universally accepted today
- Precisely defined deterministic operations
- No direct access to IEEE 754 in high-level programming languages
- Compiler writers decide how formulas are translated into IEEE 754 operations

For reproducibility, consider the compiler an integral part of your program!

The future of reproducibility

The future of reproducibility

Reproducibility will be guaranteed automatically by the computer.

Technologies for ensuring the integrity of distributed digital information:

- checksums
- content-addressed storage

Already used by:

- Git, Software Heritage
- Nix, Guix
- IPFS / IPLD

Remaining challenges:

- Reconciling interactivity with reproducibility
- Replicability
- Verifiability