BioPython and NumPy

- Specialized libraries streamline bioinformatics research.
- Libraries make advanced capabilities easier to implement.
- Do you need to “roll your own?”

UNIVERSITEIT STELLENBOSCH UNIVERSITY

BioPython

- SeqIO enables reading and writing common formats for molecular biology.
- PairwiseAligner enables alignment of sequence pairs, and Bio.Align can handle multiple sequence alignments.
- Bio.PDB can read and access protein structure data.
- In short, common bioinformatics functions are just a library call away!

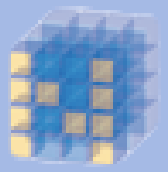
PJA Cock et al. *Bioinformatics* (2009) 25: 1422-1423.

From line of text to full sequences

```
AccessionCount = 0
LetterCounts = collections.Counter()
FASTAFile = SeqIO.parse("thing.fasta", "fasta")

try:
    for SequenceRecord in FASTAFile:
        AccessionCount += 1
        LetterCounts.update(SequenceRecord.seq)

except FileNotFoundError:
    print("Error: The file was not found!")
    quit()
```



NumPy

UNIVERSITEIT STELLENBOSCH UNIVERSITY

Library for scientific computing

- Enables modeling of N-dimensional arrays
- "Broadcasts" arithmetic operations on arrays of different shapes
- Makes complex tools like Fast Fourier Transform easily accessible
- Opens the door to many other libraries for scientific computing

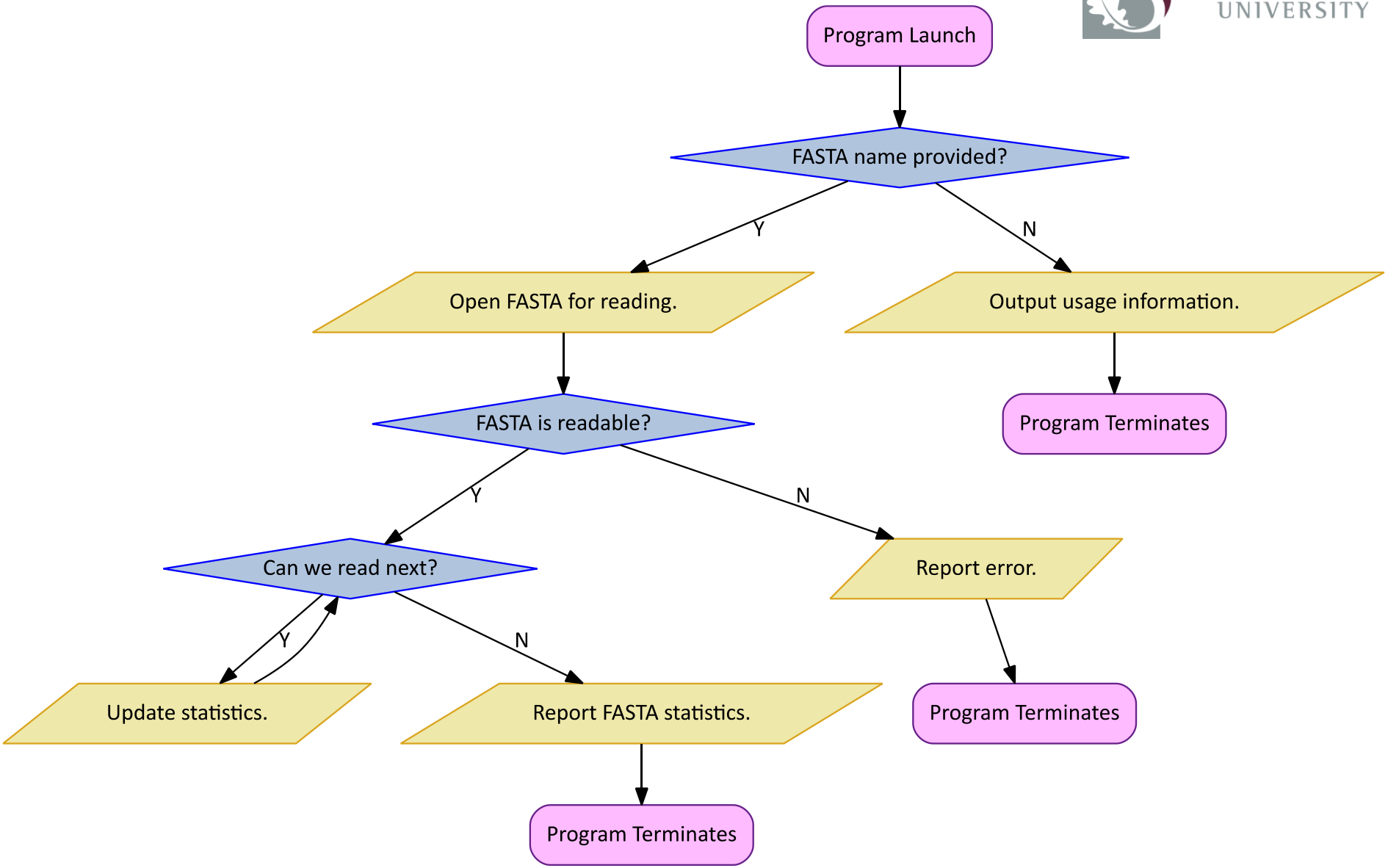
Broadcasting example

```
import numpy
A = numpy.array([[0,0,0],[10,10,10],[20,20,20]])
B = numpy.array([1,2,3])

>>> print(A)
[[ 0  0  0]
 [10 10 10]
 [20 20 20]]

>>> print(B)
[1 2 3]

>>> print(A+B)
[[ 1  2  3]
 [11 12 13]
 [21 22 23]]
```



FASTA sniffer