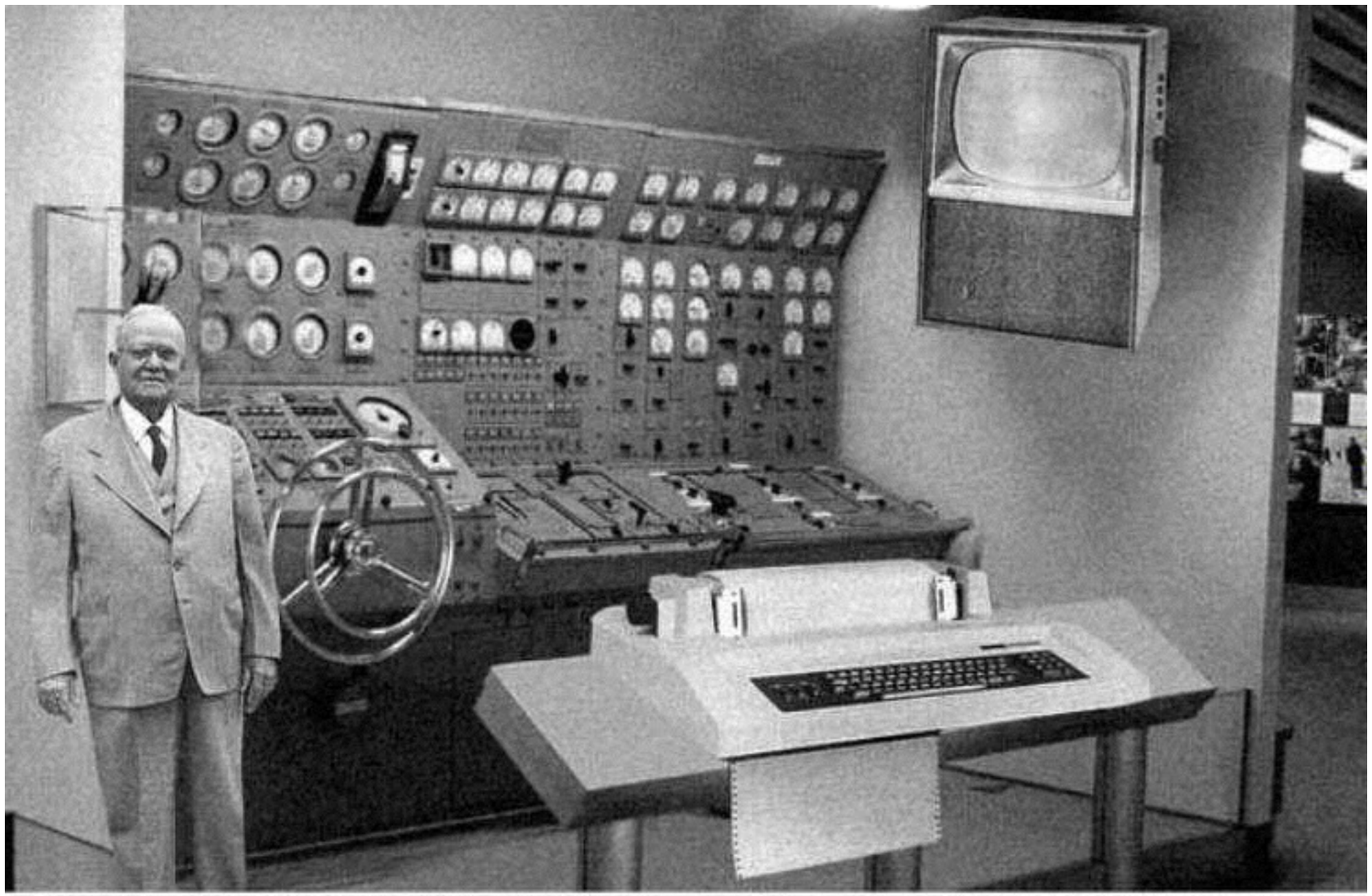
Making Better Figures

Image Formats

Matthew Kenworthy // Leiden Observatory // 2016

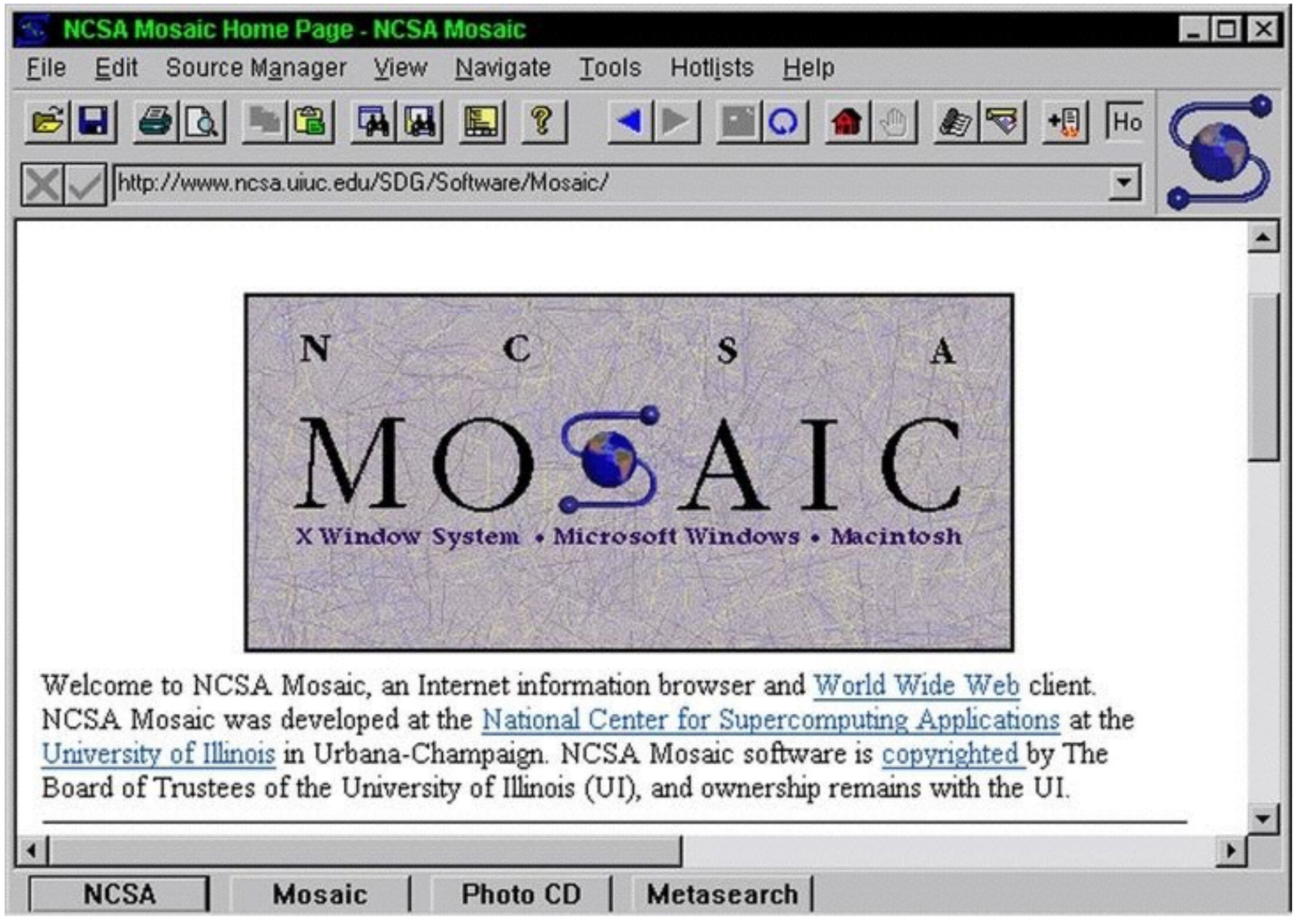
prehistory of the internet



Scientists from the RAND Corporation have created this model to illustrate how a "home computer" could look like in the year 2004. However the needed technology will not be economically feasible for the average home. Also the scientists readily admit that the computer will require not yet invented technology to actually work, but 50 years from now scientific progress is expected to solve these problems. With teletype interface and the Fortran language, the computer will be easy to use.

Matthew Kenworthy // Leiden Observatory // 2016

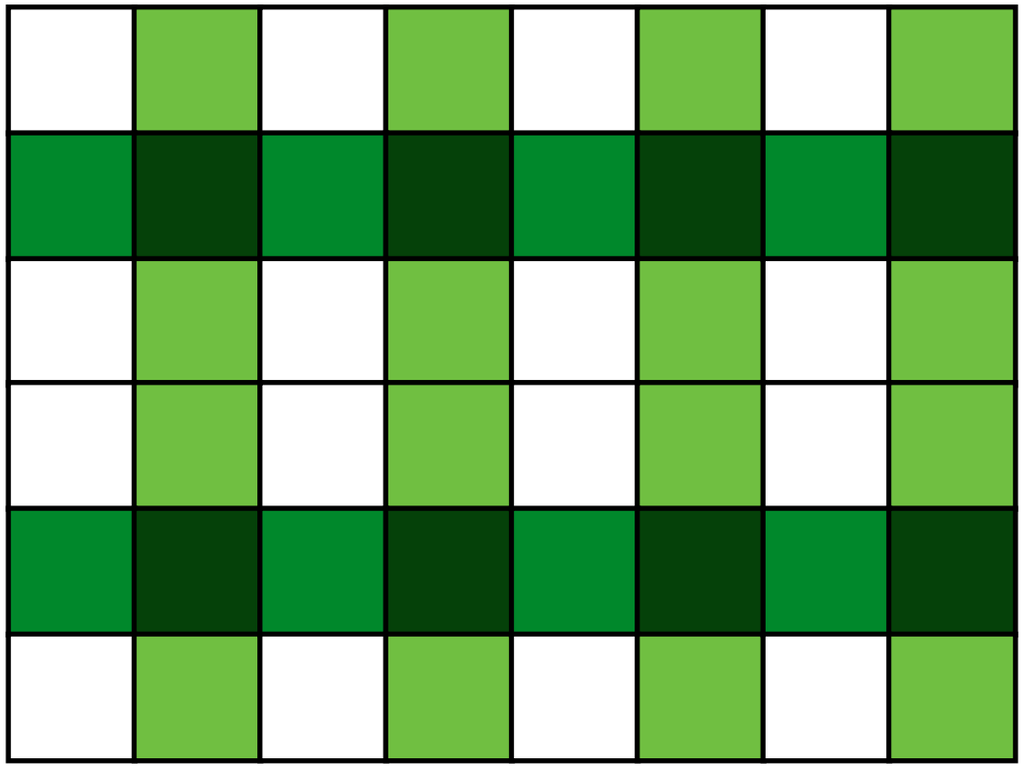
graphics limited by bandwidth and computing power



bitmap (pixel) formats

e.g. TIFF, GIF, JPG, PNG….

Images are represented with PIXELS that can have one or more colours



…

…

…

Matthew Kenworthy // Leiden Observatory // 2016

vector formats

e.g. Postscript, PDF, SVG….

Images are represented as a set of PRIMITIVE GRAPHICAL OBJECTS and they need to be DRAWN to a screen or sheet of paper

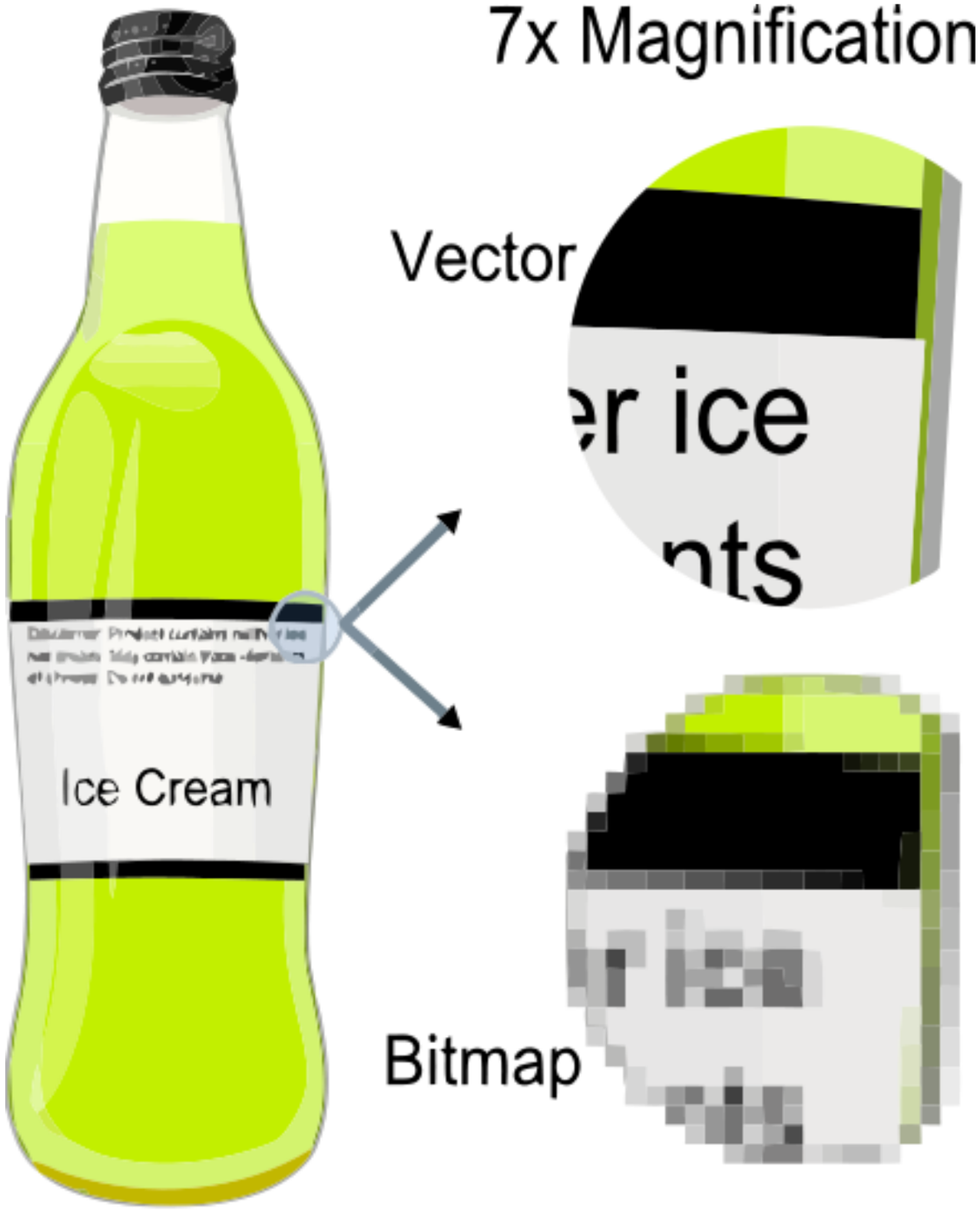
```
Start in lower left corner
move to the right 1.0 inches
Draw a blue box with a red 0.1 inch border with side 2 inches
Move to 6 inches top by 3 inches right
Write the text “Hello World!” in green
    with a point size of 72 in Helvetica
.   .   .
.   .   .
```

POINTS and NODES control the shape of LINES

Multiple connected LINES form POLYGONS with fill colours and borders

Matthew Kenworthy // Leiden Observatory // 2016

vector versus bitmap

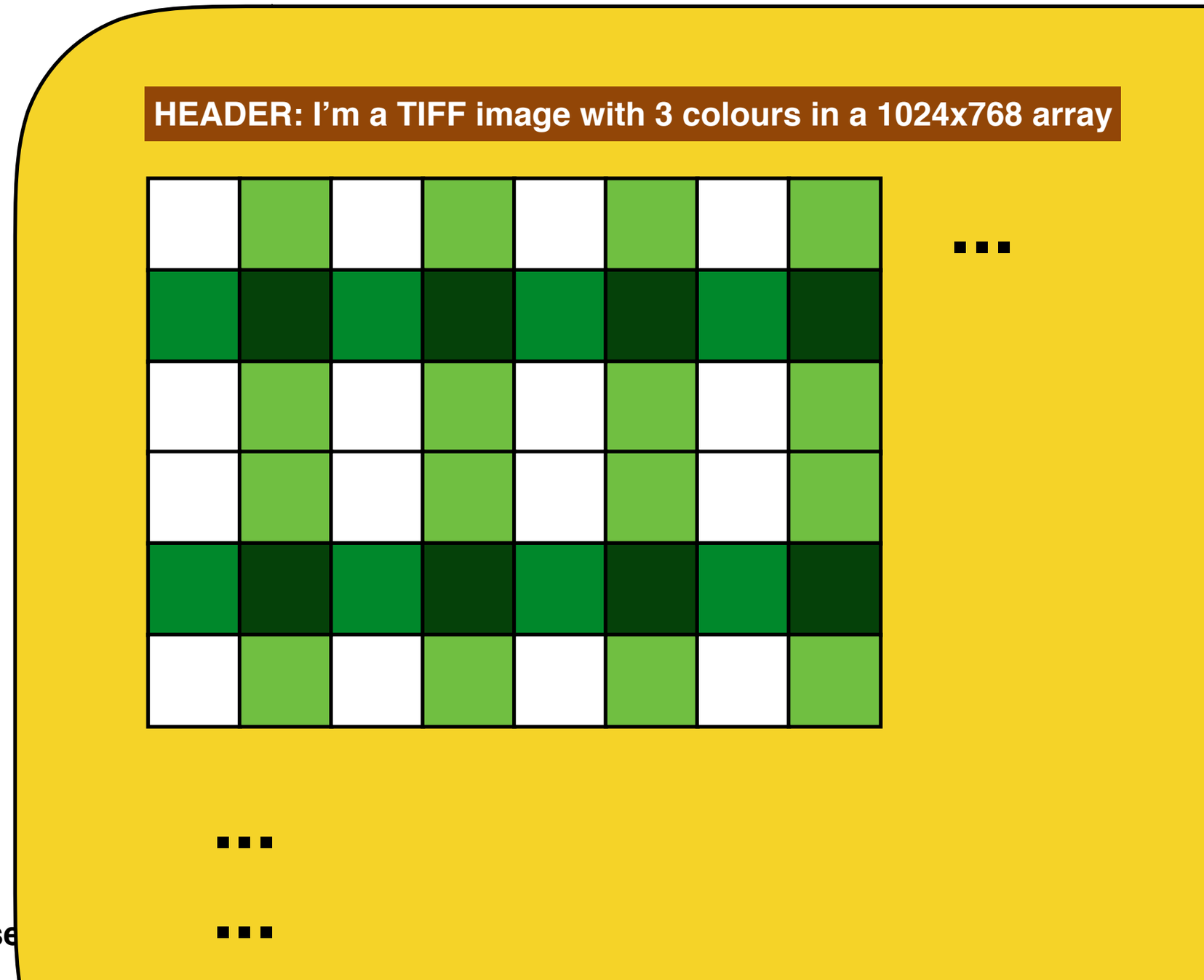


Matthew Kenworthy // Leiden Observatory // 2016

TIFF (Tagged Image File Format)

colour images stored as RED, GREEN, BLUE 2D arrays

optional lossless and lossy compression

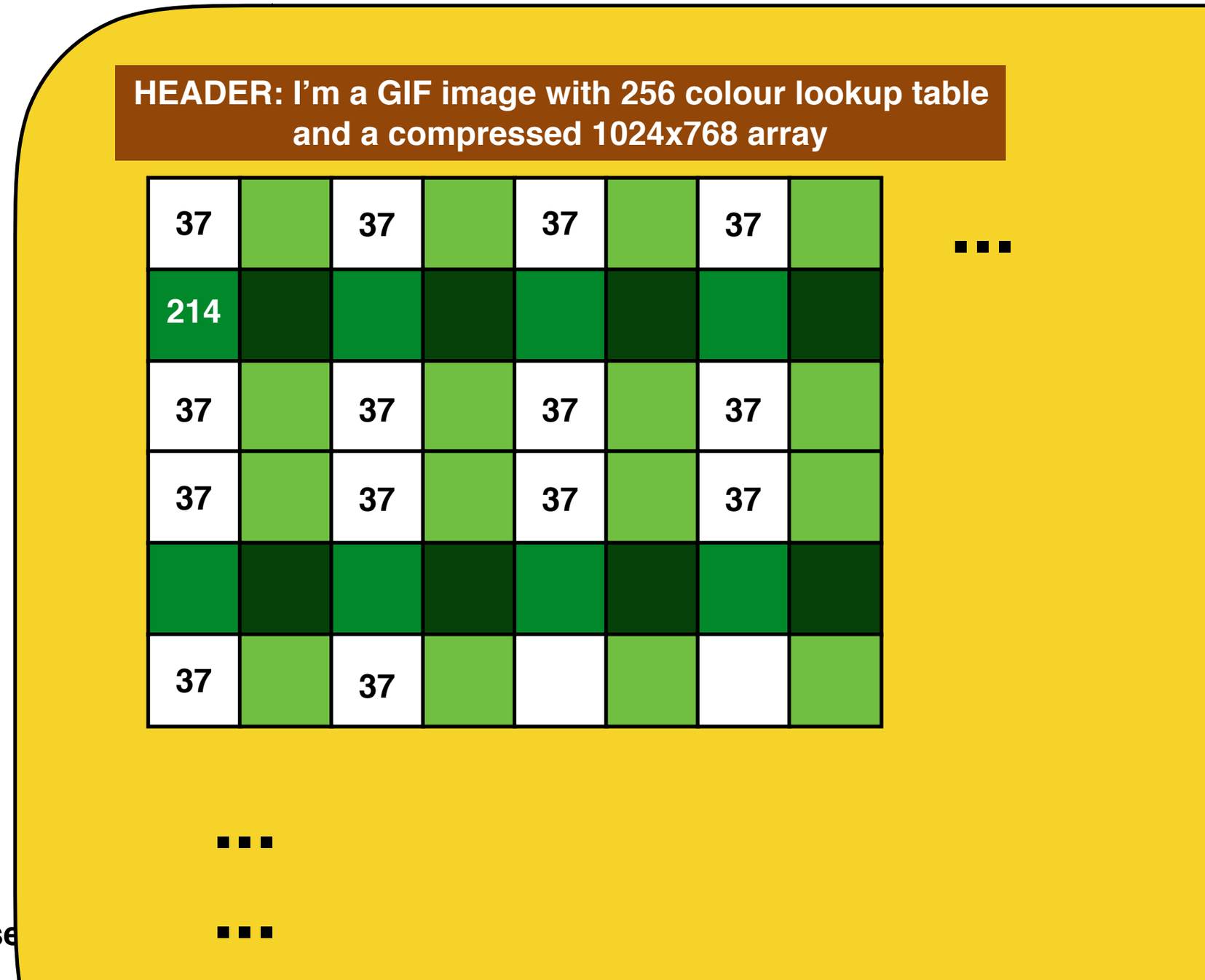


GIF (Graphics Interchange Format)

One of the earliest standards used on the internet
Optimised for low bandwidth and low colour graphics displays

colour images as 2d array of 1 byte (256 levels) per pixel with a colour look-up table

Image data is COMPRESSED losslessly



GIF also supports animations

Matthew Kenworthy // Leiden Observatory // 2016

JPEG (Joint Photographic Experts Group)

Images stored as sets of COEFFICIENTS of a DISCRETE COSINE FOURIER TRANSFORM

Smoothly varying compression across this image

Variable compression a tradeoff for file size

PNG (Portable Network Graphics)

Pixel based format

PNG handles transparent colours

Superior to JPEG in almost every way….

Matthew Kenworthy // Leiden Observatory // 2016

postscript

Vector format written in English(!)
You can edit a postscript file with a text editor
owned by Adobe corporation

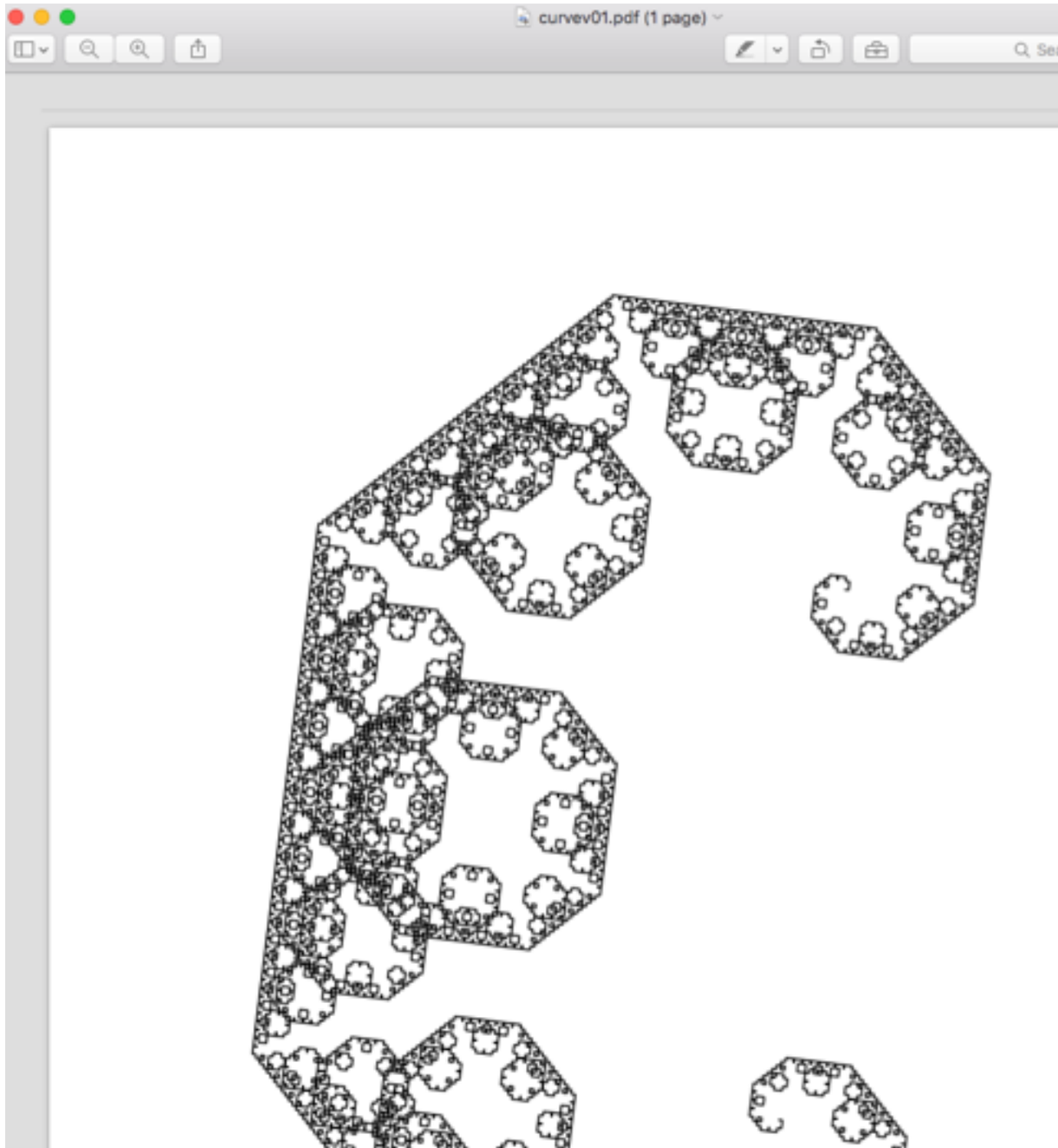
```
%!PS
 /Courier              % name the desired font
 20 selectfont         % choose the size in points and establish
                       % the font as the current one
 72 500 moveto         % position the current point at
                       % coordinates 72, 500 (the origin is at the
                       % lower-left corner of the page)
(Hello world!) show    % stroke the text in parentheses
showpage               % print all on the page
```

Matthew Kenworthy // Leiden Observatory // 2016

postscript/pdf

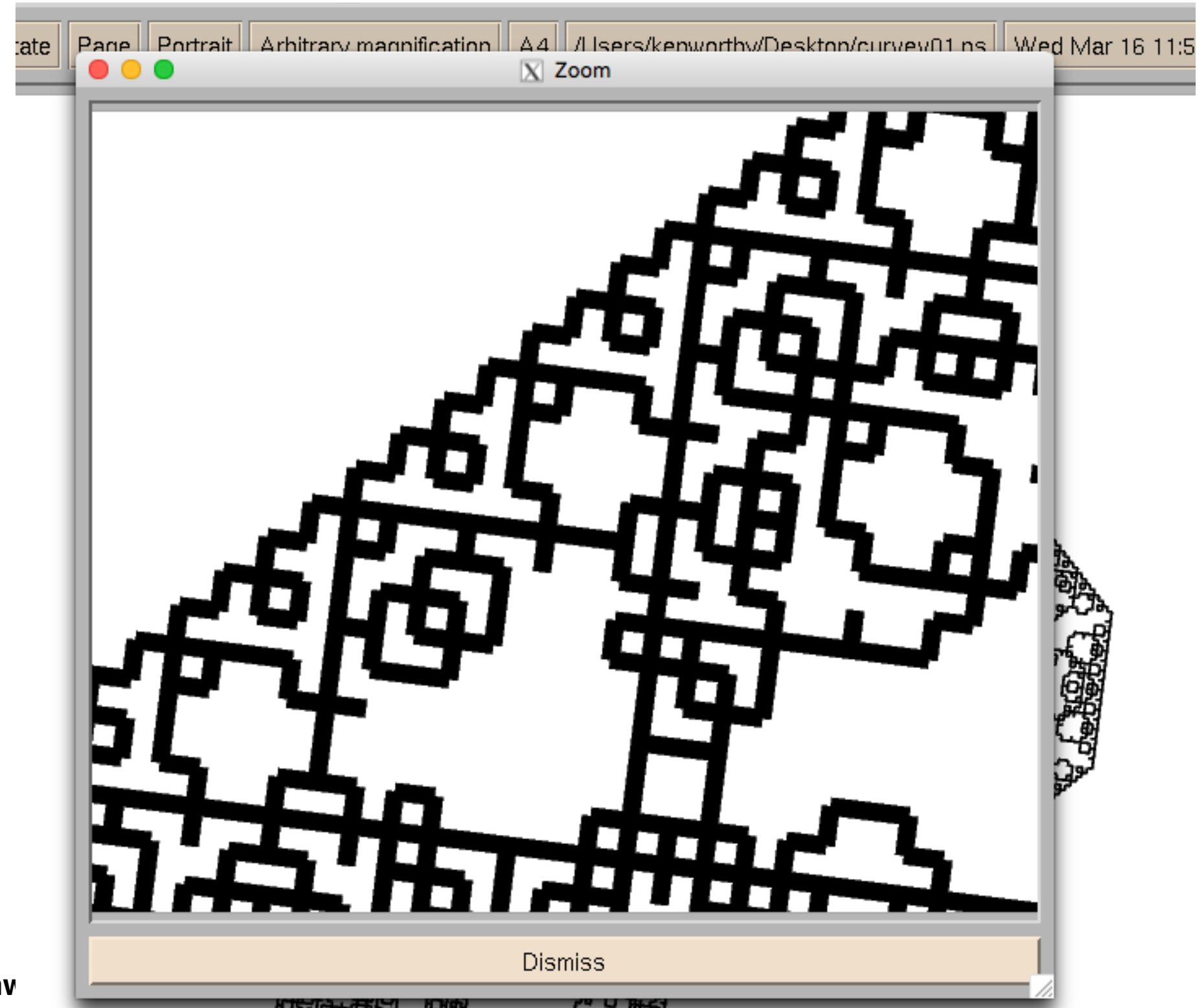
```
%!PS
/inch {72 mul} def  % define an inch in points
/draw_c_curve {
    dup
    0 gt
    {
        1 sub
        dup
        6 2 roll
        4 copy
        neg add add add 2 div
        5 copy
        pop
        add add exch neg add 2 div
        2 copy
        6 4 roll
        10 -1 roll
        draw_c_curve
        5 -1 roll
        draw_c_curve
    }
    {
        pop
        moveto
        lineto
    } ifelse
} def
newpath
5.5 inch 3.5 inch 6 inch 7.5 inch 14 draw_c_curve
stroke
showpage
```

It is also a TURING COMPLETE stack based programming language!



Matthew Kenworthy // Leiden Observatory // 2016

you can zoom in a long way



Matthew Kenw

SVG (Scalable Vector Graphics)

Most recent vector ‘standard’ and PS to SVG mostly works

Uses XML to define the graphics:

SVG code	Illustration
`<line x1="0" y1="100" x2="100" y2="0" stroke-width="2" stroke="black" />`	

Simple line

SVG code	Illustration
`<text x="15" y="45" font-size="40" fill="red">some text</text>`	some text

text

Matthew Kenworthy // Leiden Observatory // 2016

Keep in vector until you can’t

**As soon as you take a vector image and convert it to a bitmap format,
you LOSE all the information about the shapes and
you FIX the resolution of your image**

You CANNOT go back from bitmap to vector

Matthew Kenworthy // Leiden Observatory // 2016

Where vector fails

you try to render 1 million circles in your PDF image!

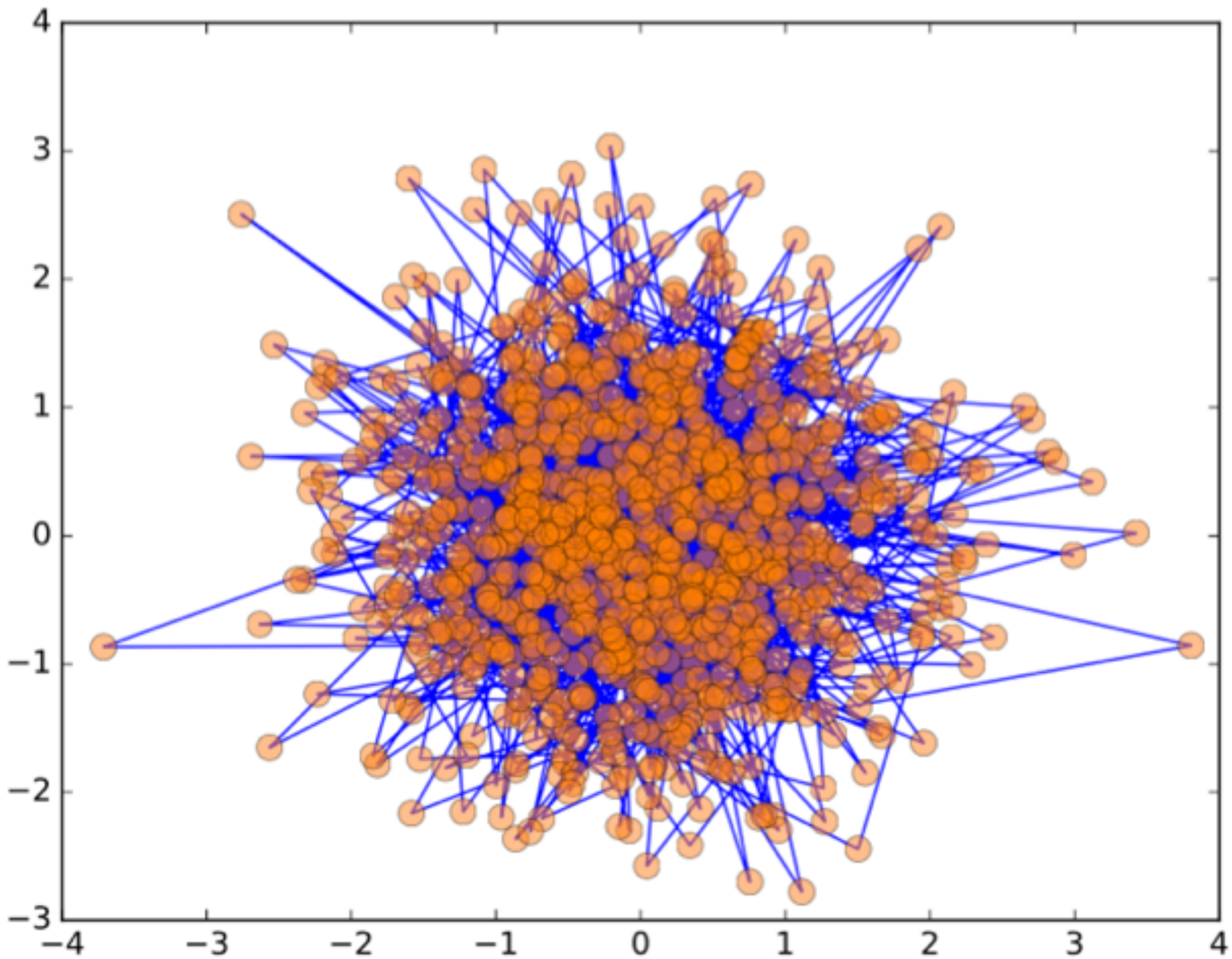
```
import matplotlib.pyplot as plt
import numpy as np

npoi = 1000000
x = np.random.randn(npoi)
y = np.random.randn(npoi)

plt.plot(x,y, 'bo-', markersize=10, markerfacecolor=(1, 0.5, 0, 0.5), )
plt.savefig('c1000000.pdf')
```

Matthew Kenworthy // Leiden Observatory // 2016

Where vector fails

you try to render 1 million circles in your PDF image!

(this took about 5 minutes to render on my computer)



Matthew Kenworthy // Leiden Observatory // 2016

Where vector fails

```
 8.8K c100.pdf
  33K c1000.pdf
 2.5M c100000.pdf
  24M c1000000.pdf
```

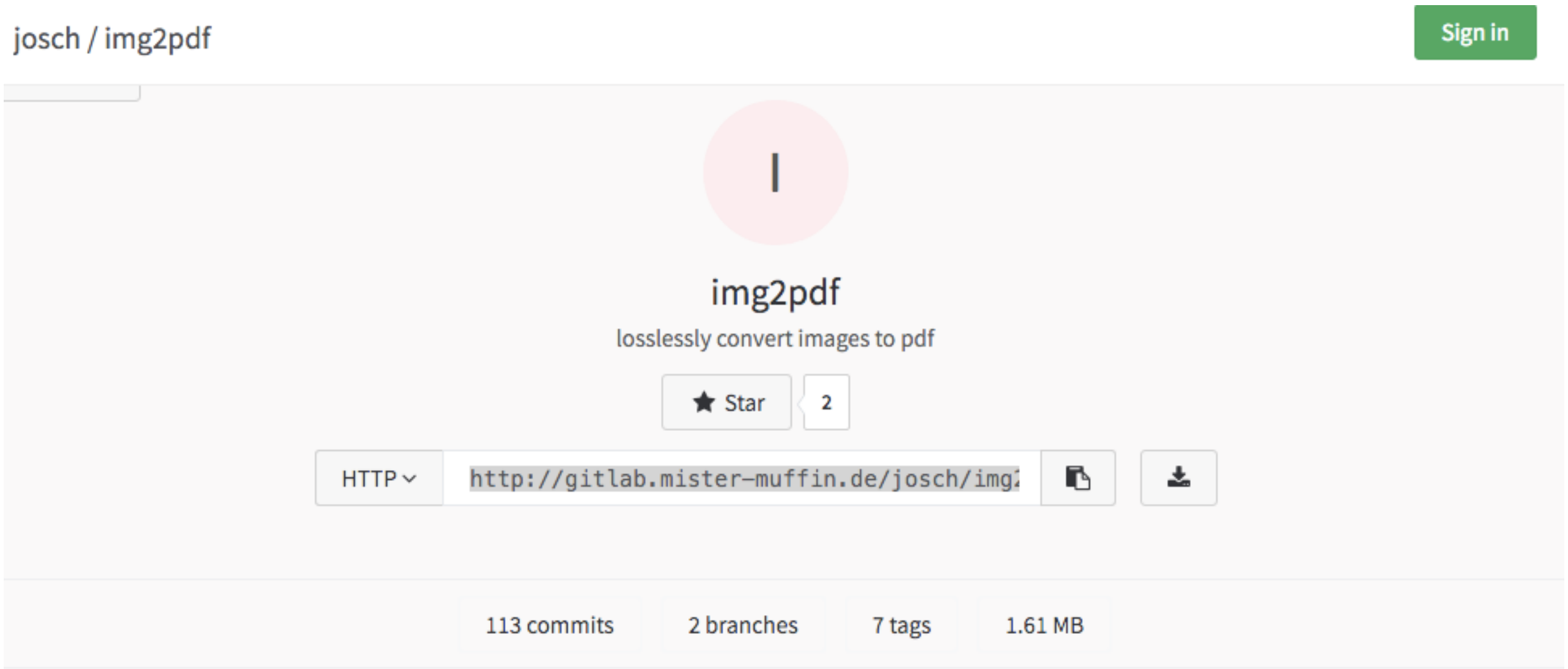
Solution: convert to high res jpg and encapsulate in PDF

```
# step 1: convert PDF to a jpg

convert -density 300 -quality 90% c1000.pdf result.jpg

# step 2: wrap jpg in PDF code with img2pdf

img2pdf --output c1000_small.pdf result.jpg

# sanity check:
ls -l result.jpg c1000_small.pdf

2217691 16 Mar 15:40 c1000_small.pdf
2216724 16 Mar 14:53 result.jpg

# …so the pdf is slightly larger than the jpg - it’s wrapped!
```

Matthew Kenworthy // Leiden Observatory // 2016

img2pdf

https://gitlab.mister-muffin.de/josch/img2pdf



c1585856 only use jp2 to parse jpeg2000 if PIL doesn't support jpeg2000 · 2016-02-17 20:31:46 +0100 by **josch**

img2pdf

Losslessly convert raster images to PDF. The file size will not unnecessarily increase. One major application would be a number of scans made in JPEG format which should now become part of a single PDF document. Existing solutions would either re-encode the input JPEG files (leading to quality loss) or store them in the zip/flate format which results into the PDF becoming unnecessarily large in terms of its file size.

jpeg2ps

Directory `tex-archive/support/jpeg2ps`

Directories

Name	Notes
dos	
os2	JPEG to PostScript converter for OS/2

Files

Name	Size	Date	Notes
Makefile	1850	1999-07-28 13:50	
asc85ec.c	3229	1999-07-28 13:50	
descrip.mms	949	1999-07-30 10:12	
getopt.c	3798	1999-07-28 13:50	
jpeg2ps-1.8.zip	298853	1999-07-30 11:20	
jpeg2ps.c	12558	1999-07-28 13:50	
jpeg2ps.dsp	2981	1999-07-30 10:14	
jpeg2ps.pdf	356766	1999-07-30 10:11	
jpeg2ps.txt	14926	1999-07-28 13:50	
psimage.h	1646	1999-07-28 13:50	
readjpeg.c	9670	1999-07-28 13:50	

jpeg2ps – Convert JPEG files to PostScript Level 2 or 3 EPS

Converts JPEG files to PostScript Level 2 or 3 EPS. In fact, jpeg2ps is not really a converter but a "wrapper": it reads the image parameters (width, height, number of color components) in a JPEG file, writes the according EPS header and then copies the compressed JPEG data to the output file. Decompression is done by the PostScript interpreter (only PostScript Level 2 and 3 interpreters support JPEG compression and decompression).

Package Details	jpeg2ps
Version	1.8
License	Do Not Sell Except by Arrangement
Copyright	1994–1999 Thomas Merz
Maintainer	Thomas Merz
Topics	prepare graphics for use with *TEX
See also	bmeps

Matthew Kenworthy // Leiden Observatory // 2016