GJØVIK UNIVERSITY COLLEGE

Course Introduction

Kjell Are Refsvik

August 24, 2011

IMT4892 Digital Workflow

http://www.ansatt.hig.no/kjellr/imt4892

Welcome!

:-)

WHERE WE ARE NOW

WEEK	TOPIC
33	-
34	Course Introduction
35	A Brief History of Computing
36	Devices and data formats, wrappers and compressors
37	Tools for transforming common datatypes
38	Moving data, remote processing and workflow automation
39	Workflow automation
40	Intellectual property rights (IPR), including Creative Commons
41	Mashup fundamentals and applied mashup using KML and Google Maps
42	XML Fundamentals, part 1
43	XML Fundamentals, part 2
44	Academic planning and writing
45	Project, week 1 - project outline delivery
46	Project, week 2 - work
47	Project delivery/presentation
48	Reading week before the exam
49	Exam

TODAY'S GOAL

- What are the course methods and means?
- What is the course goal?
- What [tech.] resources do you need to get hold of to begin this course?

ME

1992-1993	Research assistant, IT [satellite imagery, programming, UNIX]	hiof.no
1993-1994	IT-Consultant [comms., www, mac, email]	hil.no/it
1994-2008	Multimedia Producer / Project Manager	hil.no/sell
2005-2008	Regional IT-Advisor, primary education	uninettabc.no
2006-2009	M.Sc. [Information Technology - Designing Digital Environments]	hiof.no
2009-	Assistant Professor [Media tech., Media design and Media production]	hig.no/imt

YOU?

Medina, Victor	MMT-CIMET2010-Høst	victor.medina@hig.no
Mehmood, Muhammad Owais	MMT-CIMET2010-Høst	muhammad.mehmood@hig.no
Neverova, Natalia	MMT-CIMET2010-Høst	natalia.neverova@hig.no
Oncu Feier, Alexandra Ioana	MMT-CIMET2010-Høst	alexandra.feier@hig.no
Azarijafari, Parinaz	MMT2011-Høst	parinaz.azarijafari@hig.no
Liu, Xinwei	MMT2011-Høst	xinwei.liu@hig.no
Yushtina, Anna	MMT2011-Høst	anna.yushtina@hig.no
Hrdy, Ondrej	UTVEKSL-M2011-Høst	ondrej.hrdy@hig.no

YOU

- Name?
- Where do you come from?
- Got your own laptop/desktop?
- Background in computing
- What’s in your toolbox (os, tools, devices)?
- Expectations - hig | master thesis | this course
- Plans for the future?

METHODS AND STYLE

- Distribution outside and inside fronter (accessibility)
- Record and podcast lecture recordings (accessibility)
- Open systems, tools, standards and formats (accessibility)
- Lectures: Would like seminars rather than monologues

LECTURES

Title page

Course map

Learning outcome

Practical messages

Lecture slides

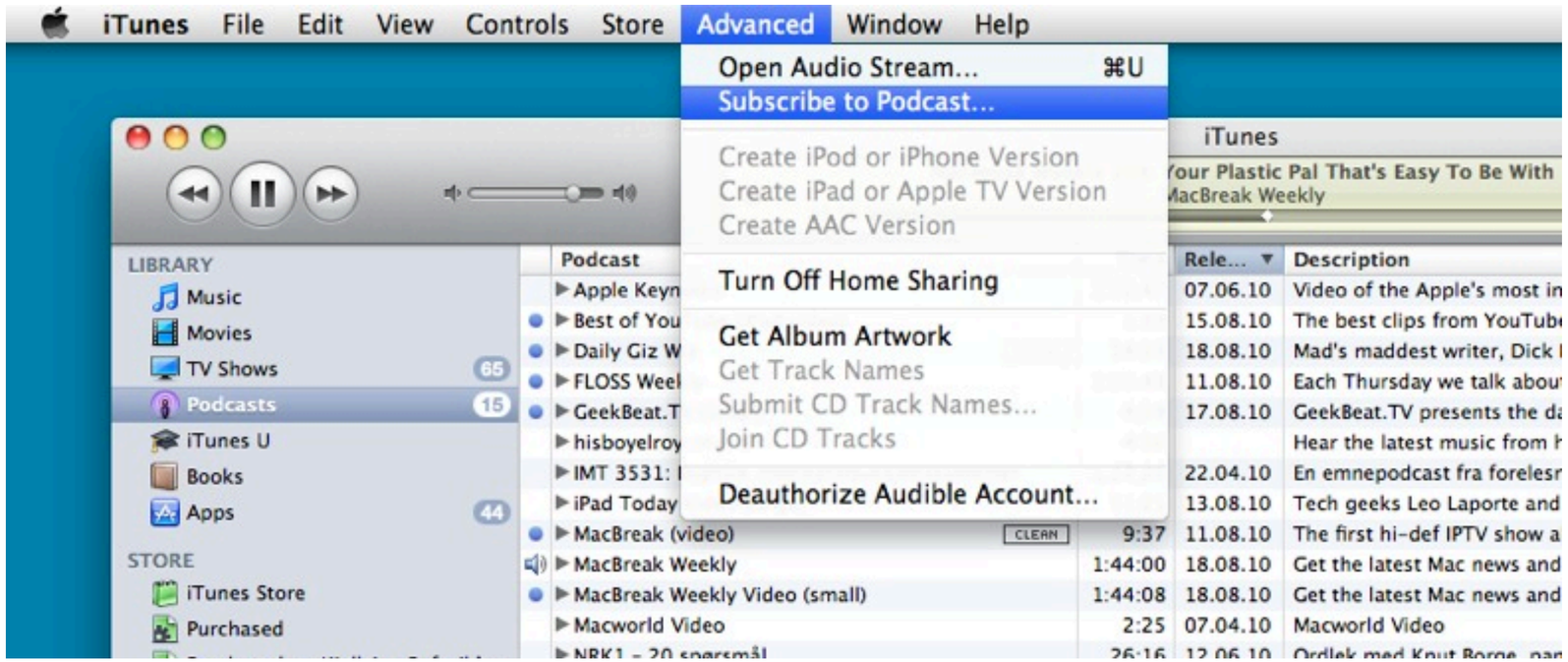
To read or do

Contact information

4892.rss

LECTURE SUBSCRIPTION AND PLAYBACK

- RSS reader (Google Reader, iTunes, ...)
- Files can also be copied to your mobile device and watched there if they support open standards including MPEG4/AAC/H.264



COURSE CONTENT

XML

IPR

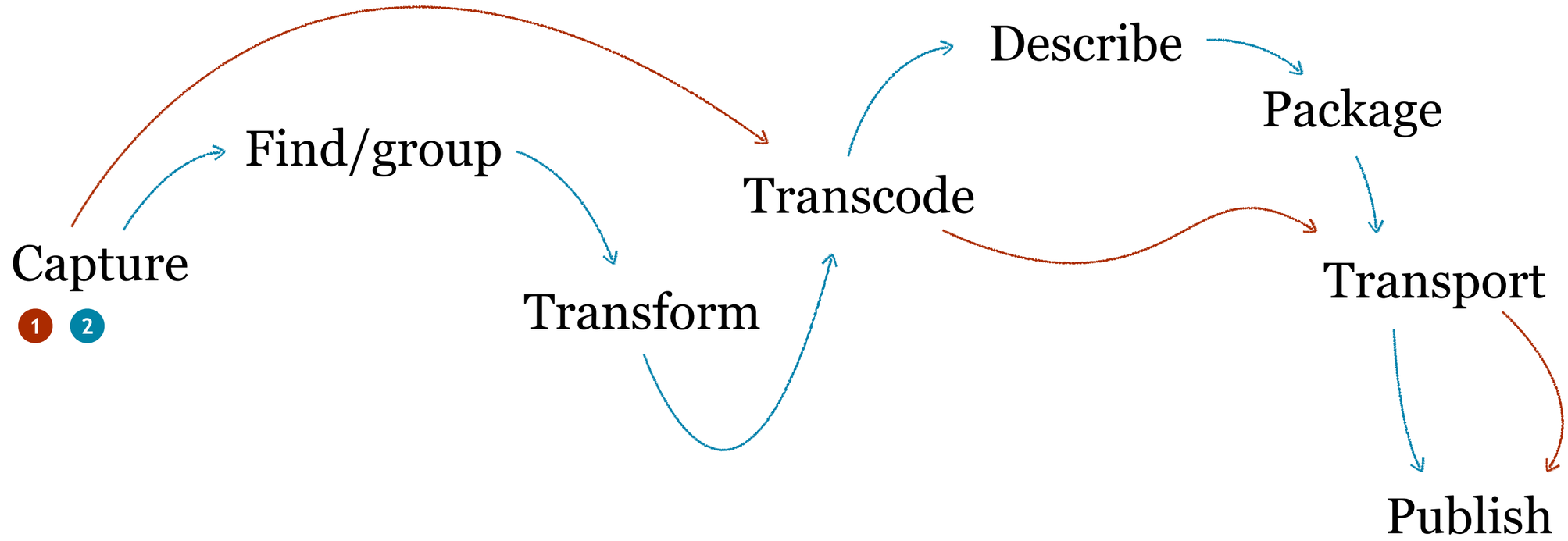
Formalising and implementing digital workflows with UNIX and open source tools

Metadata

Computer History and UNIX

Academic writing using LaTex and BibTex

IMAGES


Capture
1
2
Find/group
Transform
Transcode
Describe
Package
Transport
Publish

WORKFLOW

"A workflow consists of a sequence of connected steps. It is a depiction of a sequence of operations, declared as work of a person, a group of persons,[1] an organization of staff, or one or more simple or complex mechanisms.

Workflow may be seen as any abstraction of real work. For control purposes, workflow may be a view on real work under a chosen aspect,[2] thus serving as a virtual representation of actual work. The flow being described may refer to a document or product that is being transferred from one step to another."

- Wikipedia, Aug. 2011

WORKFLOW - IMPORTANT QUESTIONS

- What systems and programs do we use as digital workers?
- Do the systems and programs we use lend themselves well to all contexts?
- Are the tools and methods we use sustainable?
- Do we select tools that work well together?
- Do we have to re-invest in learning new tools
- Investing [heavily] in learning tools/systems now that you may benefit from later?

SCENARIO

“These 8000 images needs a custom watermark, a size reduction and needs to be online at this URL before lunch...”

LAB WORK

- I will be there to ask tech questions
- You may send me mail with tech questions as well
- Also - help eachother with tech issues. The one helping ends up learning just as much as the one getting help.

LAB WORK

- I will be there to ask tech questions
- You may send me mail with tech questions as well
- Also - help eachother with tech issues. The one helping ends up learning just as much as the one getting help.

ASSIGNMENTS

- I will give small assignments (not mandatory)
- I will ask you to show off your work in front of the other students
  - Training your presentation skills
  - Pushing you to deliver
- We will base most of our examples and small exercises around color images

TOOLBOX

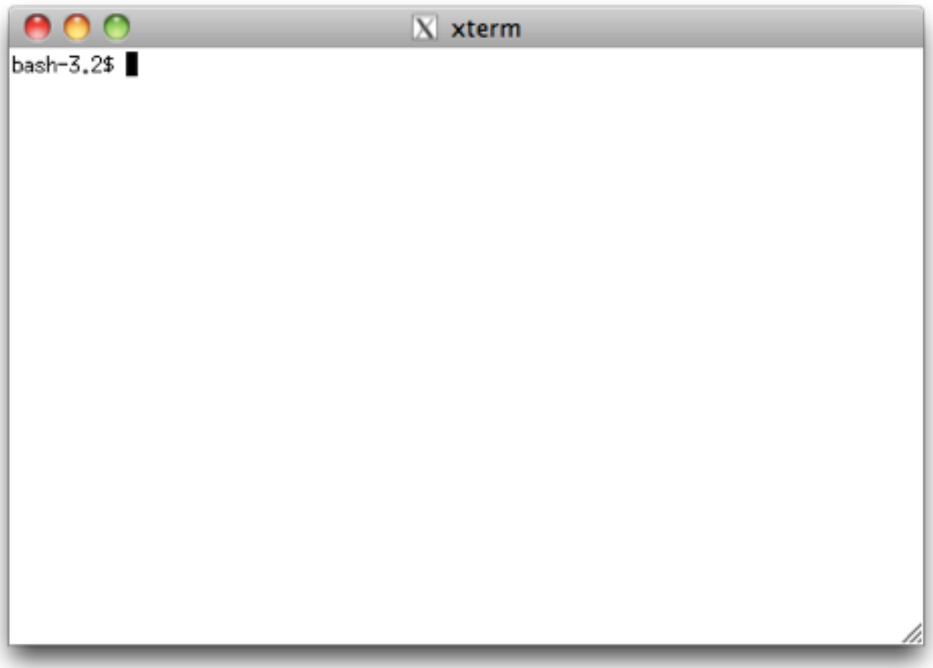
imagemagick

ffmpeg

gnuplot

gpsbabel

exiftool



sed

awk

grep

find

latex

ssh

scp

bibtex

sftp

cp

mv

cut

PROJECT AT THE END OF THIS COURSE

- Mini project where you need to outline what you want, and then try to deliver
- Will include writing your (perhaps?) first referenced academic document
- We suggest the use of LaTex/BibTex and a good editor to do the work

UNIX?

- 1.Your Own laptop running
  - PC: UBUNTU
  - Mac: Mac OS X BSD
- 2. Access stan.hig.no through a terminal window (i.e. PuTTY)
- 3.Borrow a Mac in A211 Mac lab

SETTING UP UNIX ON YOUR MACHINES

- Helping each other out using Social Media
  - Twitter: #imt4892 (open, fast, unencumbered)
  - G+?
- Because by helping others you will help yourself learn better and more

INFORMATION

- Elect class representatives (MMT and CIMET) and report names to Rune Hjelsvold (Vice Dean). Rune.hjelsvold@hig.no
  - Not much work
  - Great learning experience

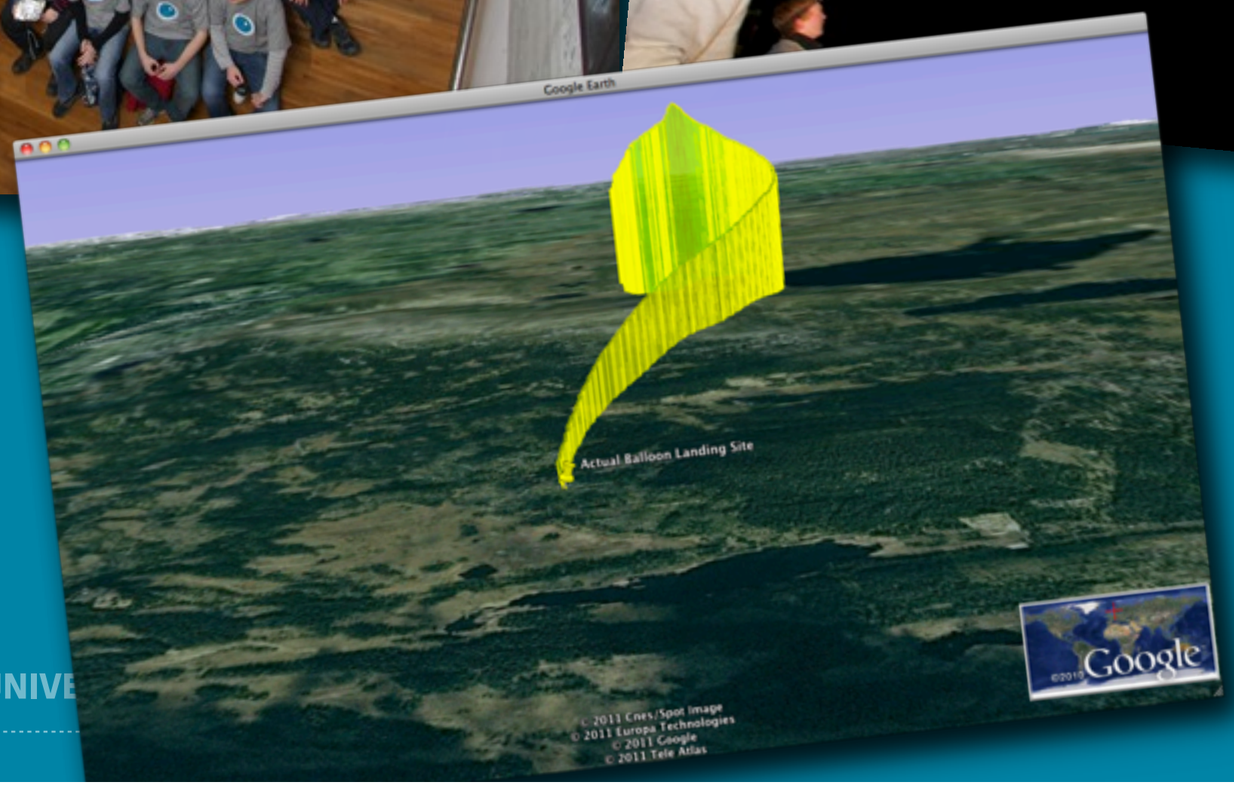
ASSIGNMENTS

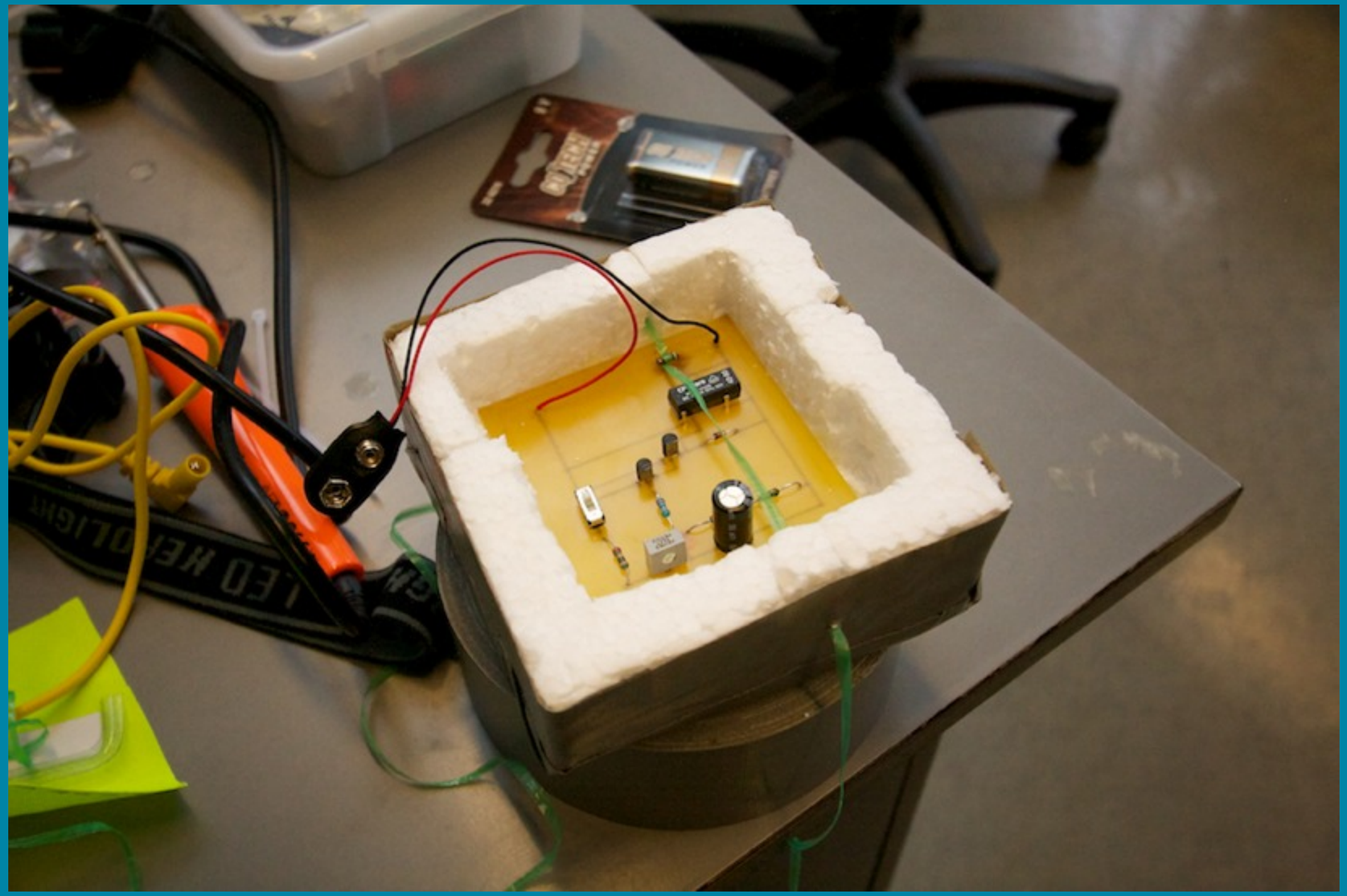
1. Buy the books
2. Start reading the first chapter in all of the books
3. Get access to a UNIX system
4. Start thinking about a project to do later in this course
5. Get hold of an RSS-reader and subscribe to the lecture feed: (feed://www.ansatt.hig.no/kjellr/imt4892/imt4892.rss)
6. E-mail me your student number so that we can generate UNIX logins for you on stan.hig.no
7. Form a social network around the course and help eachother!

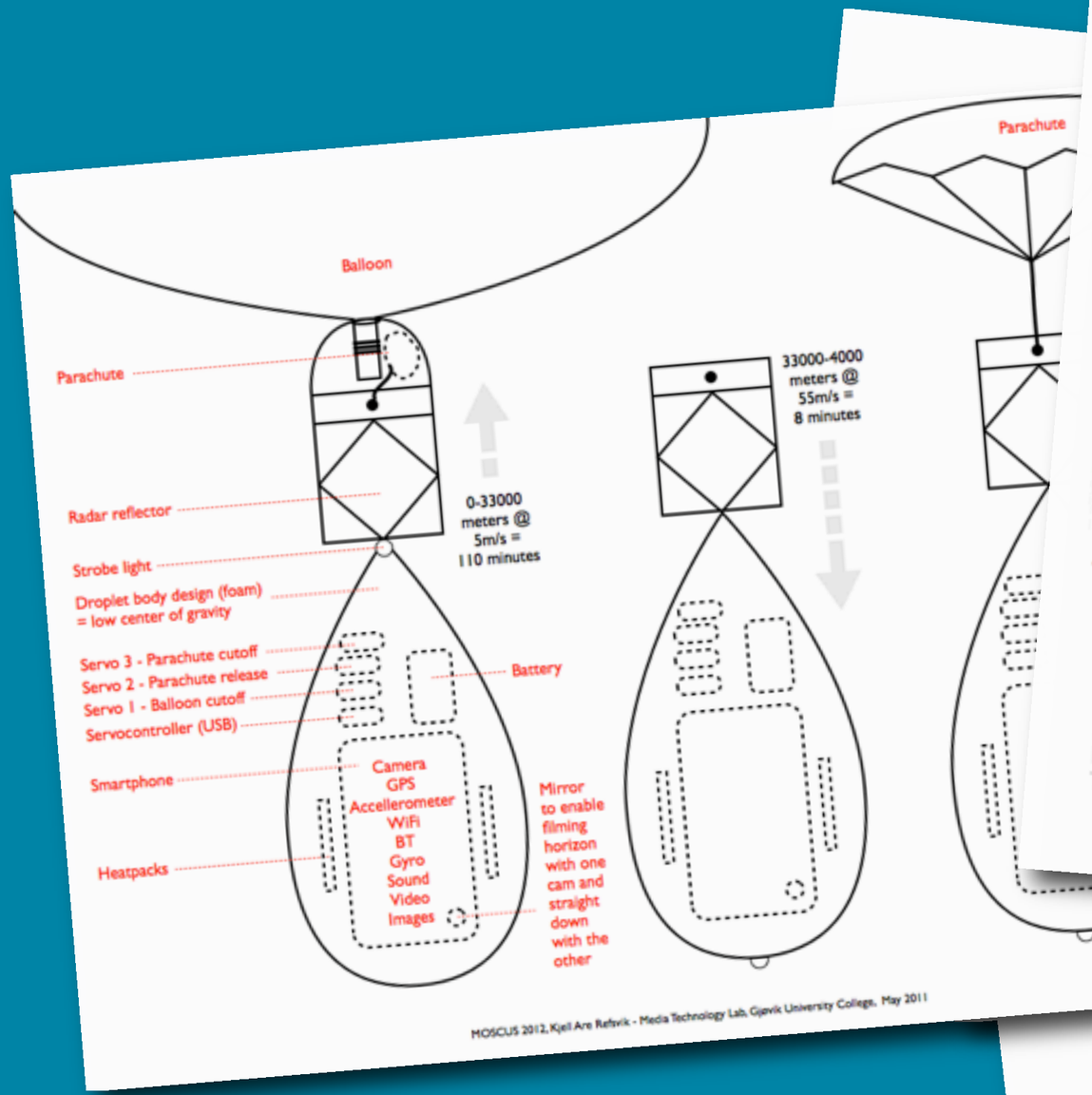
ONE MORE THING...

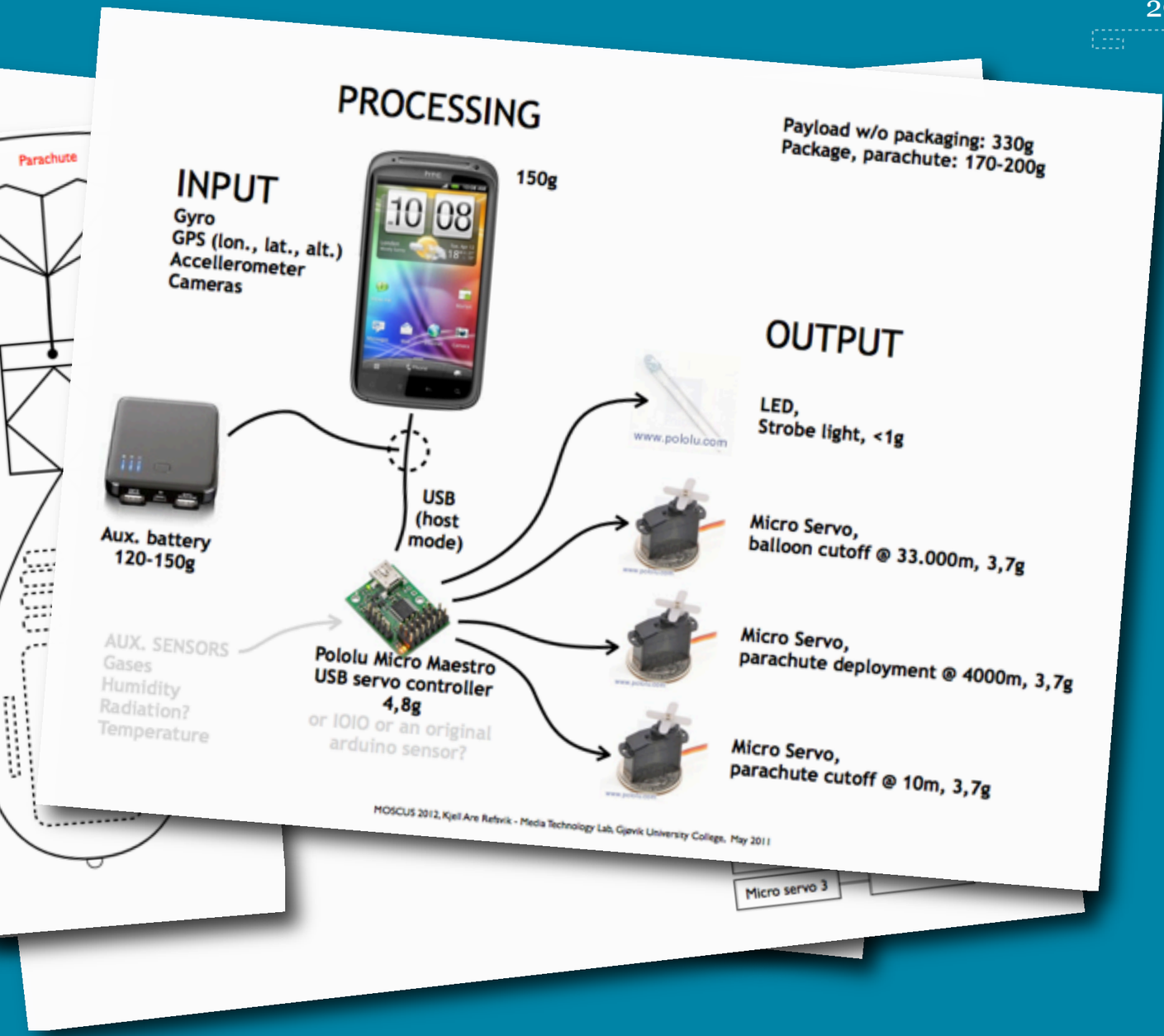
Mobile Solutions and Data Collection in the Upper Atmosphere

GJØVIK UNIVERSITY COLLEGE


Google Earth
Actual Balloon Landing Site
© 2011 Cnes/Spot Image
© 2011 Europa Technologies
© 2011 Google
© 2011 Tele Atlas
Google


LED HEADLIGHT


Balloon
Parachute
Radar reflector
Strobe light
Droplet body design (foam)
= low center of gravity
Servo 3 - Parachute cutoff
Servo 2 - Parachute release
Servo 1 - Balloon cutoff
Servocontroller (USB)
Smartphone
Heatpacks
Camera
GPS
Accellerometer
WiFi
BT
Gyro
Sound
Video
Images
Battery
Mirror
to enable
filming
horizon
with one
cam and
straight
down
with the
other
0-33000
meters @
5m/s =
110 minutes
33000-4000
meters @
55m/s =
8 minutes
Parachute
MOSCUS 2012, Kjell Are Refsvik - Media Technology Lab, Gjøvik University College, May 2011


PROCESSING
INPUT
Gyro
GPS (lon., lat., alt.)
Accellerometer
Cameras
150g
Payload w/o packaging: 330g
Package, parachute: 170-200g
OUTPUT
Aux. battery
120-150g
USB
(host
mode)
LED,
Strobe light, <1g
Micro Servo,
balloon cutoff @ 33.000m, 3,7g
Micro Servo,
parachute deployment @ 4000m, 3,7g
Micro Servo,
parachute cutoff @ 10m, 3,7g
AUX. SENSORS
Gases
Humidity
Radiation?
Temperature
Pololu Micro Maestro
USB servo controller
4,8g
or IOIO or an original
arduino sensor?
Micro servo 3
MOSCUS 2012, Kjell Are Refsvik - Media Technology Lab, Gjøvik University College, May 2011

MOSCUS 2012

- Need a group of students interested in mobile phones, RC toys and programming for an extra curriculum activity
- I have parts, a smartphone, experiences from 2011 and plans for 2012
- Task: build and test a hardware platform for a near space instrument package based around an Android phone.

http://www.ansatt.hig.no/kjellr/projects/moscus/

CONTACT INFORMATION

Name	Kjell Are Refsvik
E-mail	kjell.refsvik@hig.no
Telephone	61 13 52 28 | 405 50 454
WWW	http://www.ansatt.hig.no/kjellr
Office	A224A-B
Available	https://hevn.hig.no/#user=kjell.refsvik@hig.no&view=cal

GJØVIK UNIVERSITY COLLEGE

A Brief History of Computing and an Introduction to Unix

Kjell Are Refsvik

August 31, 2011

IMT4892 Digital Workflow

http://www.ansatt.hig.no/kjellr/imt4892

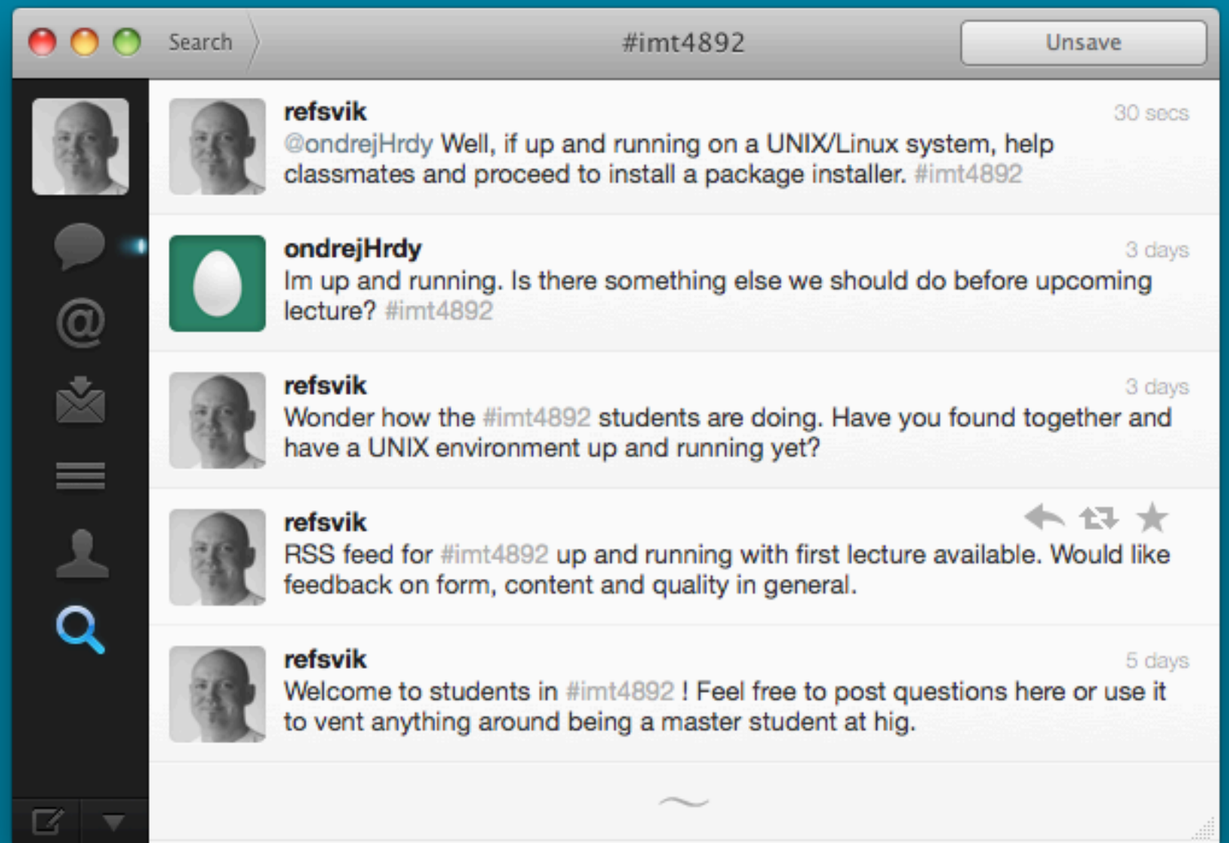
20 WHERE WE ARE NOW

WEEK	TOPIC
33	-
34	Course Introduction
35	A Brief History of Computing and an Introduction to Unix
36	Devices and data formats, wrappers and compressors
37	Tools for transforming common datatypes
38	Moving data, remote processing and workflow automation
39	Workflow automation
40	Intellectual property rights (IPR), including Creative Commons
41	Mashup fundamentals and applied mashup using KML and Google Maps
42	XML Fundamentals, part 1
43	XML Fundamentals, part 2
44	Academic planning and writing
45	Project, week 1 - project outline delivery
46	Project, week 2 - work
47	Project delivery/presentation
48	Reading week before the exam
49	Exam

ASSIGNMENTS FROM LAST WEEK?

1. Buy the books
2. Start reading the first chapter in all of the books
3. Get access to a UNIX system
4. Start thinking about a project to do later in this course
5. Get hold of an RSS-reader and subscribe to the lecture feed:
   (feed://www.ansatt.hig.no/kjellr/imt4892/imt4892.rss)
6. Form a social network around the course and help eachother!

TWITTER, YESTERDAY AT LUNCH



STAN.HIG.NO

- you now all have accounts on stan.hig.no
- `ssh stan.hig.no` in a terminal window [ssh=secure shell]
- username: s...
- password: p...
- Change password at first login with `passwd`

TODAY'S GOAL

- Be able to identify the key milestones in the development and use of computers
- Know the two main user interface paradigms and their strengths and weaknesses
- Know the main attributes of a UNIX command-line system and its strength as an environment for digital workflows

TOOLS

- No content without form - no form/design without tools
- The tools we select influence how we work
- Ideally, we should have different tools and interface paradigms in our toolbox to handle different situation
- Approaches may range from having an artistic relationship with data, to having a more mechanical one

COMPUTER?

- Minimum: Input, processing, output
- Mechanical as well as electrical/electronical
- 300 BC - Mechanical devices for calculation (i.e. Abacus)
- 1801 - Mechanical calculation machines (i.e. Looms)
- 1884 - Electrical computing with vacuum tube
- 1925 - Transistor
- 1949 - Intergrated circuit
- 1971 - Microprosessor

Sector, England, c. 1740
Loan of Gwen and Gordon Bell (B338.85)

Sector, France, c. 1680
Loan of Gwen and Gordon Bell,

Sector, Italy, 1687
Loan of Gwen and Gordon Bell (B1507.01)

Mechanical calculation device, 1687

Mechanical computing, 1847-1849 / 2008

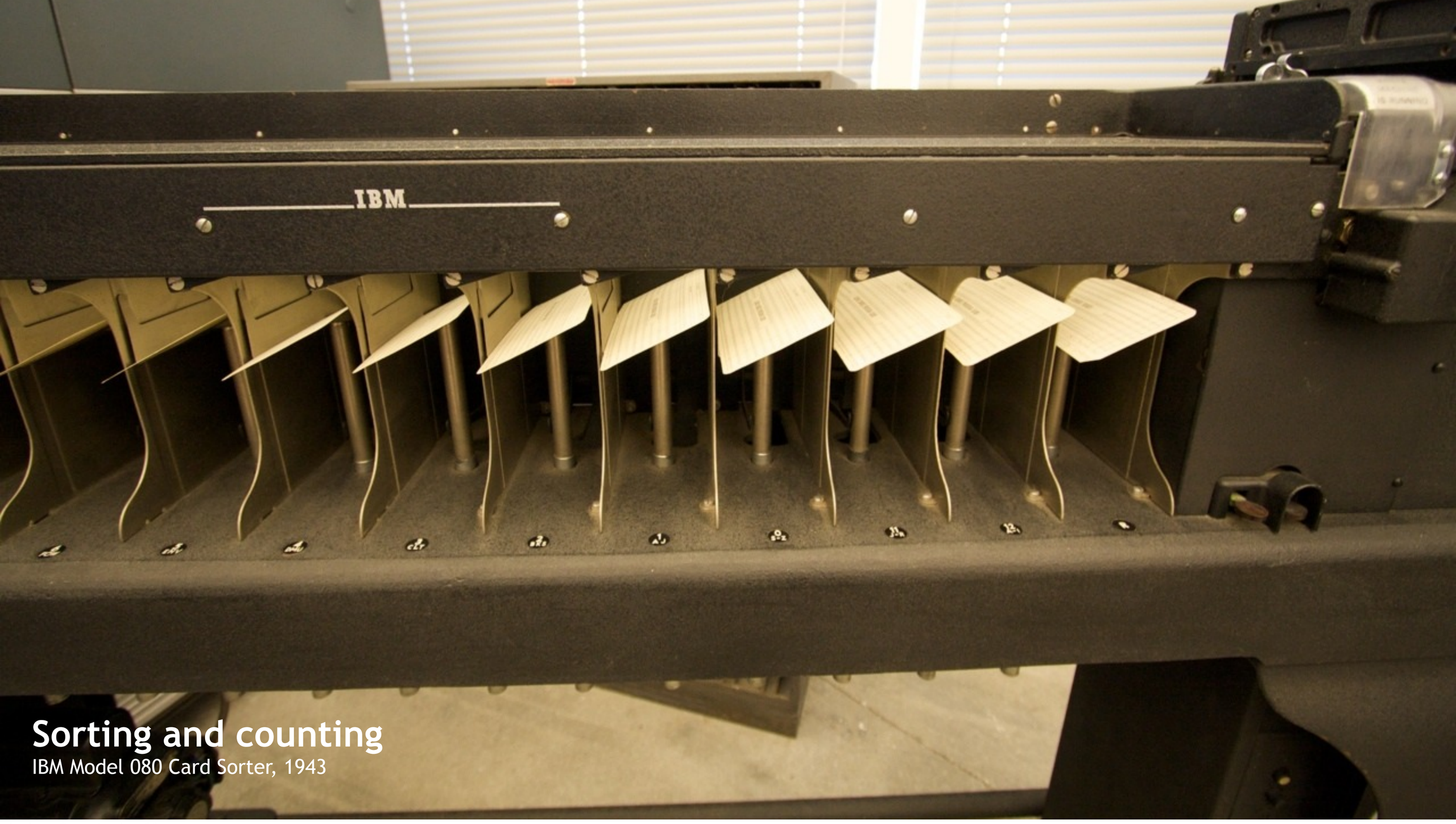
The difference Engine no. 2, Charles Babbage

Mechanical computing, 1847-1849 / 2008

The difference Engine no. 2, Charles Babbage

Mechanical computing, 1847-1849 / 2008

The difference Engine no. 2, Charles Babbage



Sorting and counting

IBM Model 080 Card Sorter, 1943

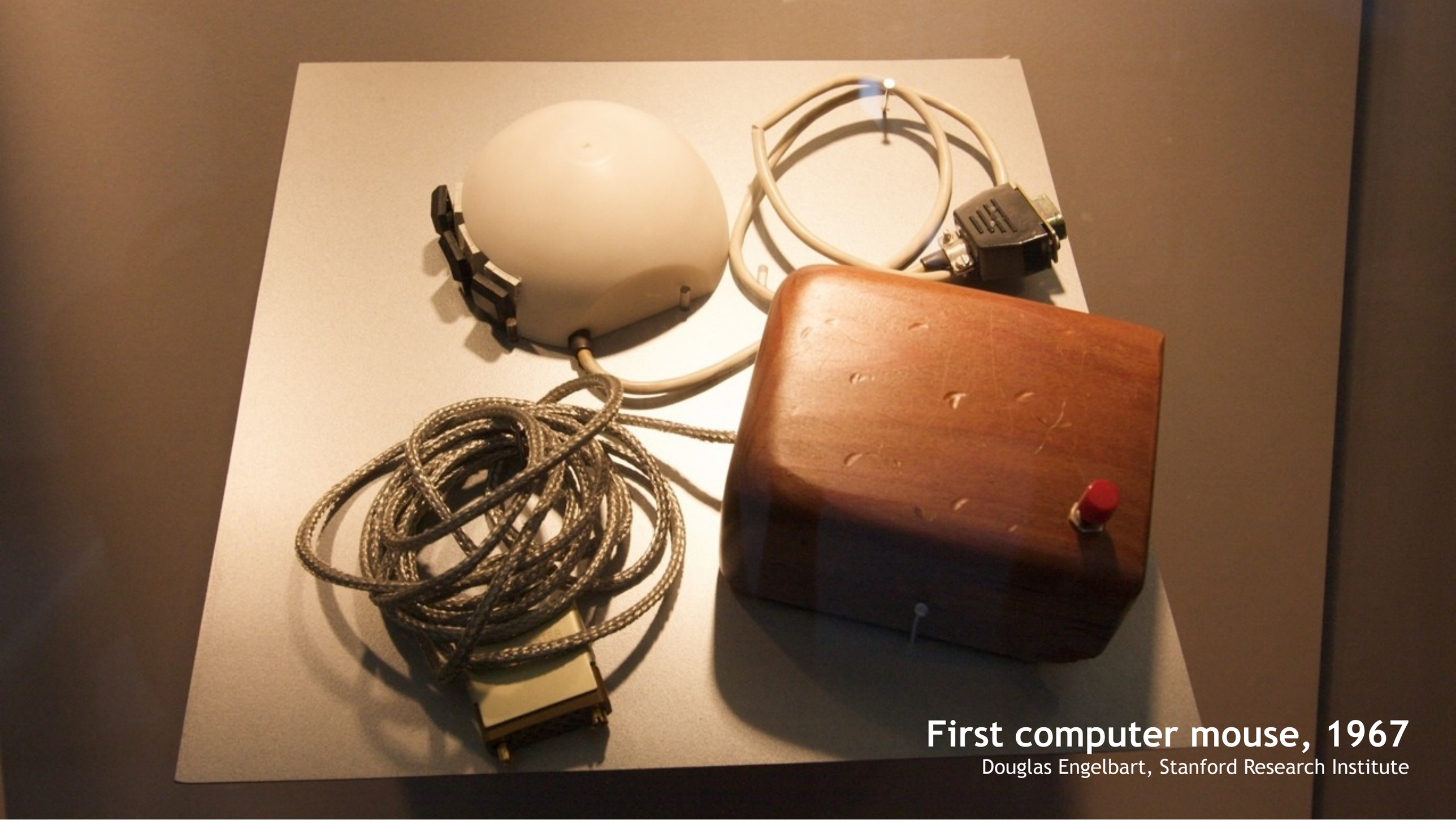
1940s and -50s. Ballistics.

Analog Computing

General Precision Systems, ca. 1950

1960s. The era of the Mainframes

Lawrence Livermore National Laboratory



First computer mouse, 1967

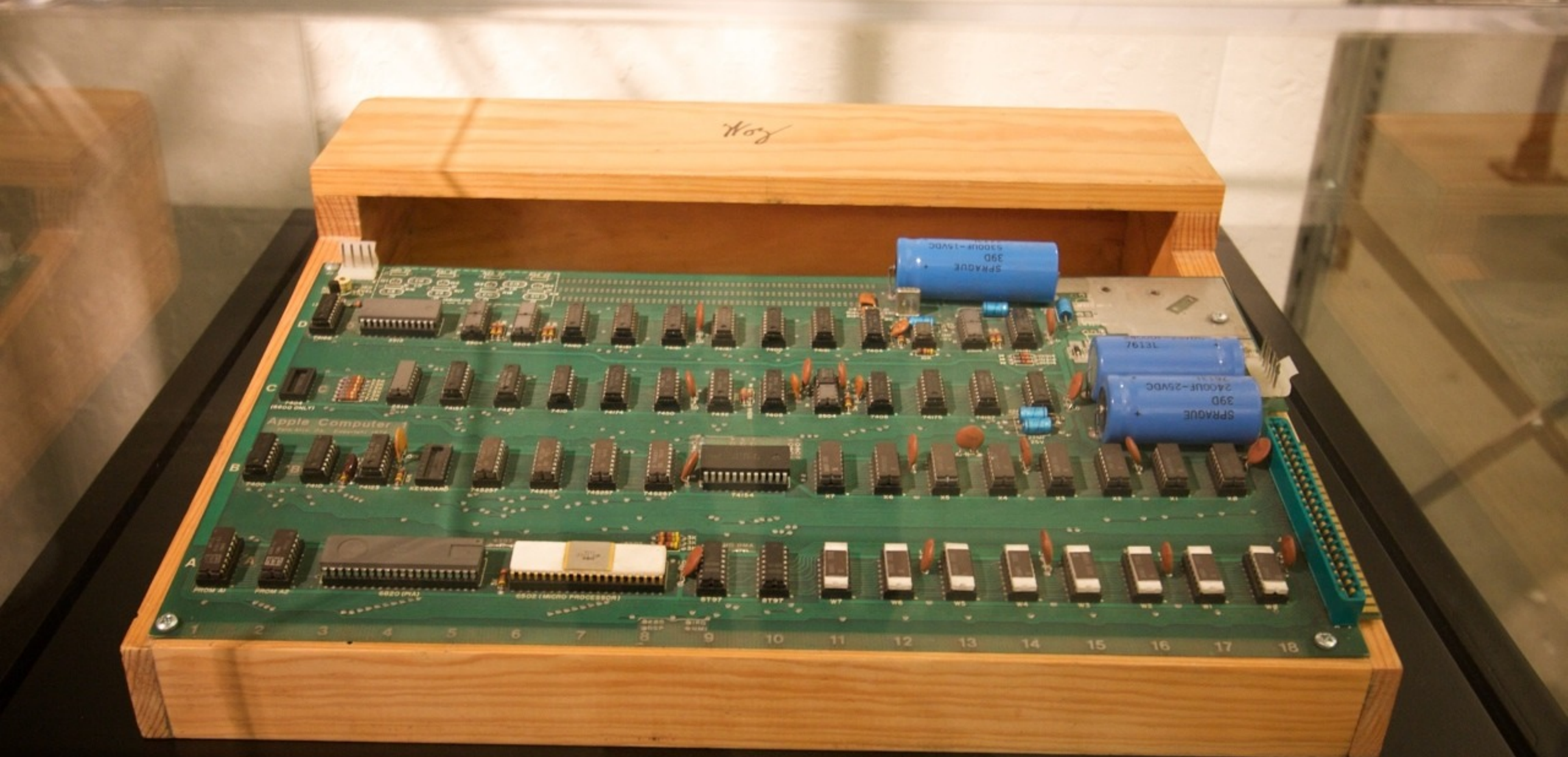
Douglas Engelbart, Stanford Research Institute

http://www.youtube.com/watch?v=JfIgzSoTMOs

http://en.wikipedia.org/wiki/The_Mother_of_All_Demos

Hobby Computing, 1975

MITS



Personal computer (Apple-1), 1976

Apple Computer, Steve Wozniac and Steve Jobs

Personal Computer, 1981

IBM PC

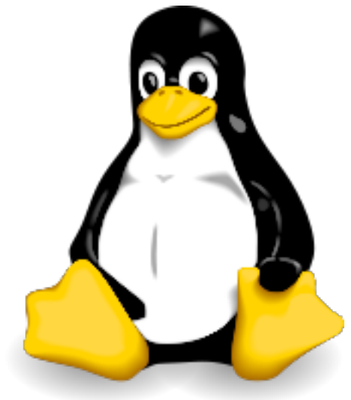
Home computing, 1982-1994

Commodore 64, Commodore Inc.

Mobile computing, 1990s

Apple Newton Messagepad, 1993, Apple Computer

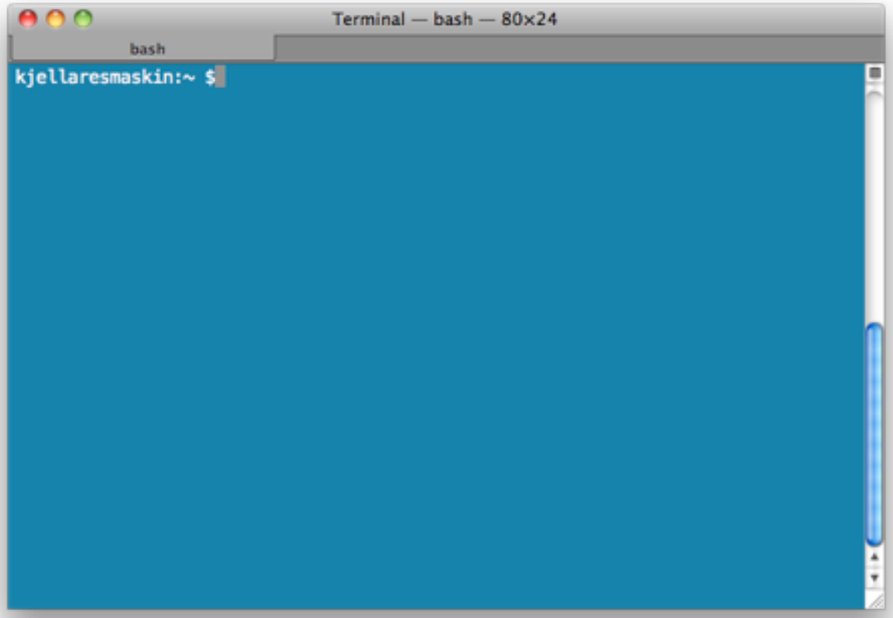
1991

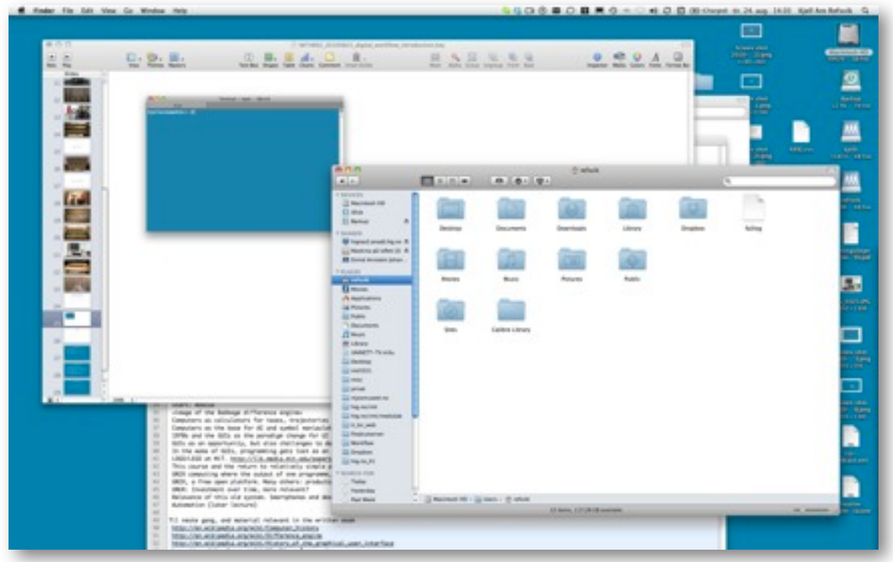


http://en.wikipedia.org/wiki/Linux

Switches, dials and levers

Paper/cardboard





TUI/CLI

GUI

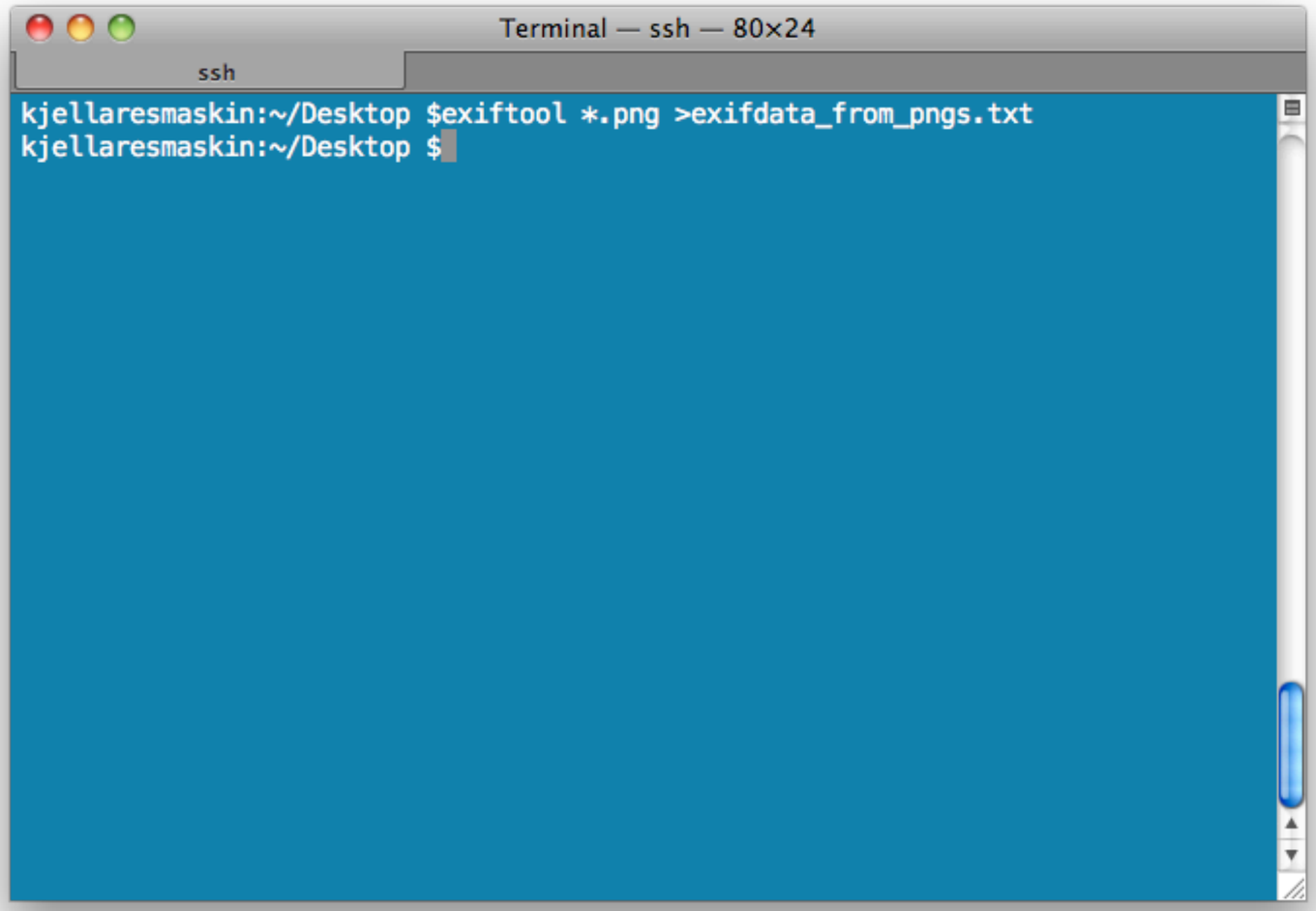
UNIX SYSTEMS, SHELLS AND COMMAND-LINE INTERFACES

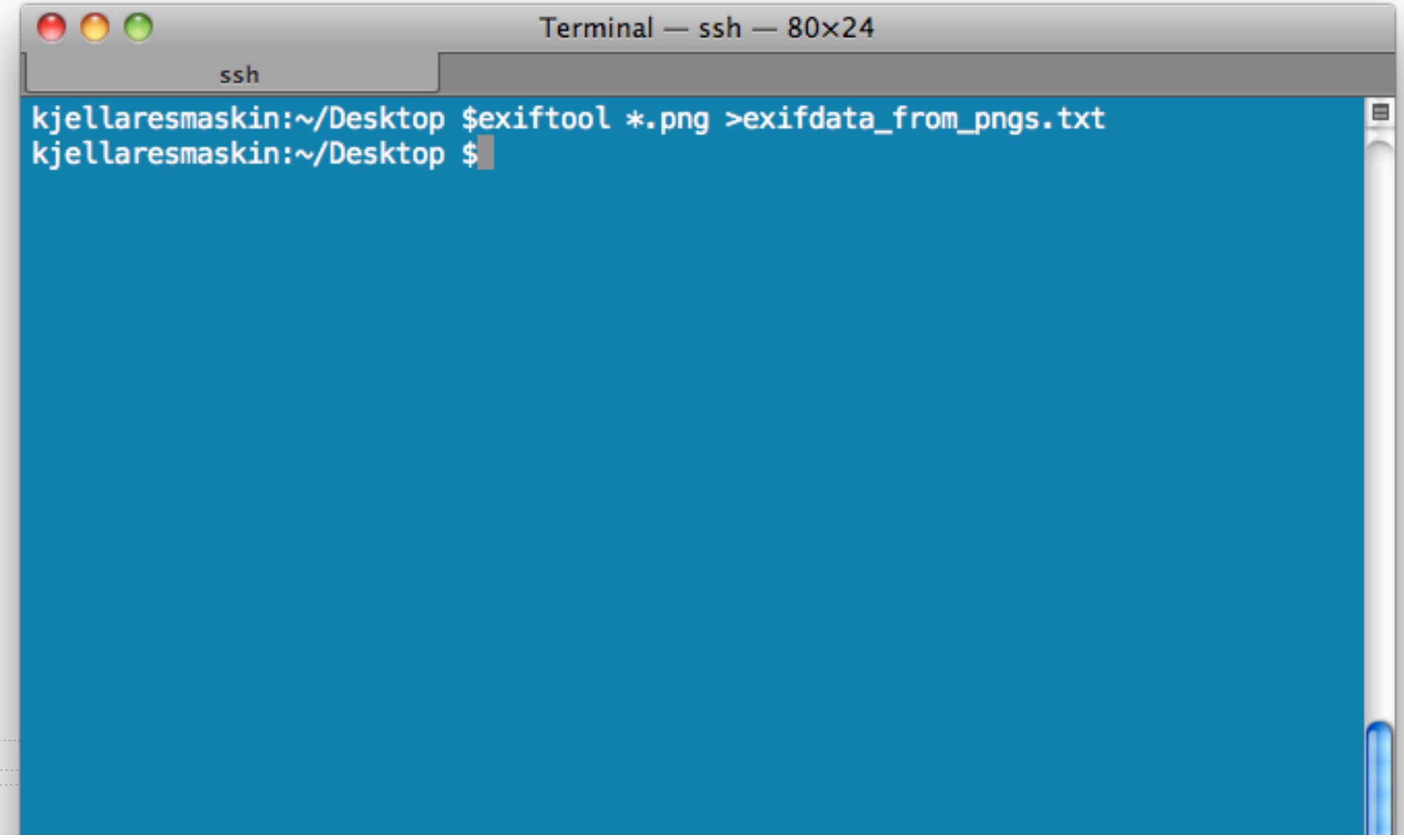
- Open and free environment where you may formalise workflows
- Small simple programs that can communicate with eachother
- Lends itself well to large and/or repetitive tasks
- Avoids lock-ins between tools and data
- As soon as the workflow had been defined, often more effective (cpu/power/time)
- Not intuitive
- Not artistic

COMMAND-LINE PROGRAMS

- `[doSomething] [how] [toFiles]`
- `[doSomething] [how] [sourceFile] [destinationFile]`
- `[doSomething] [how] < [inputFile] > [outputFile]`
- `[doSomething] [how] | [doSomething] [how] | [do Something] [how] > [outputFile]`

http://en.wikipedia.org/wiki/Command_line_argument#Anatomy_of_a_shell_CLI


Terminal — ssh — 80×24
ssh
kjellaresmaskin:~/Desktop $exiftool *.png >exifdata_from_pngs.txt
kjellaresmaskin:~/Desktop $

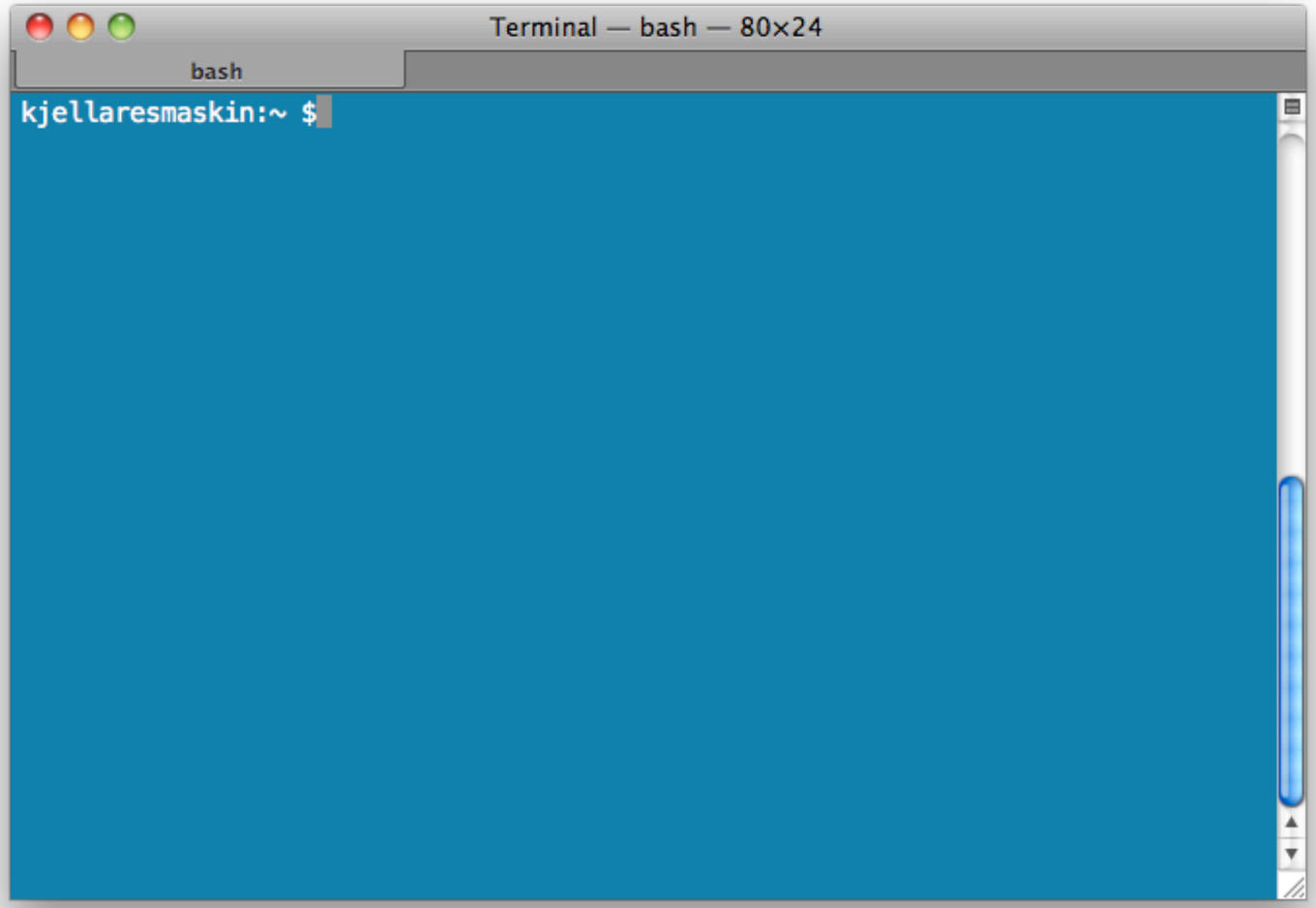

Terminal — ssh — 80×24
ssh
kjellaresmaskin:~/Desktop $exiftool *.png >exifdata_from_pngs.txt
kjellaresmaskin:~/Desktop $

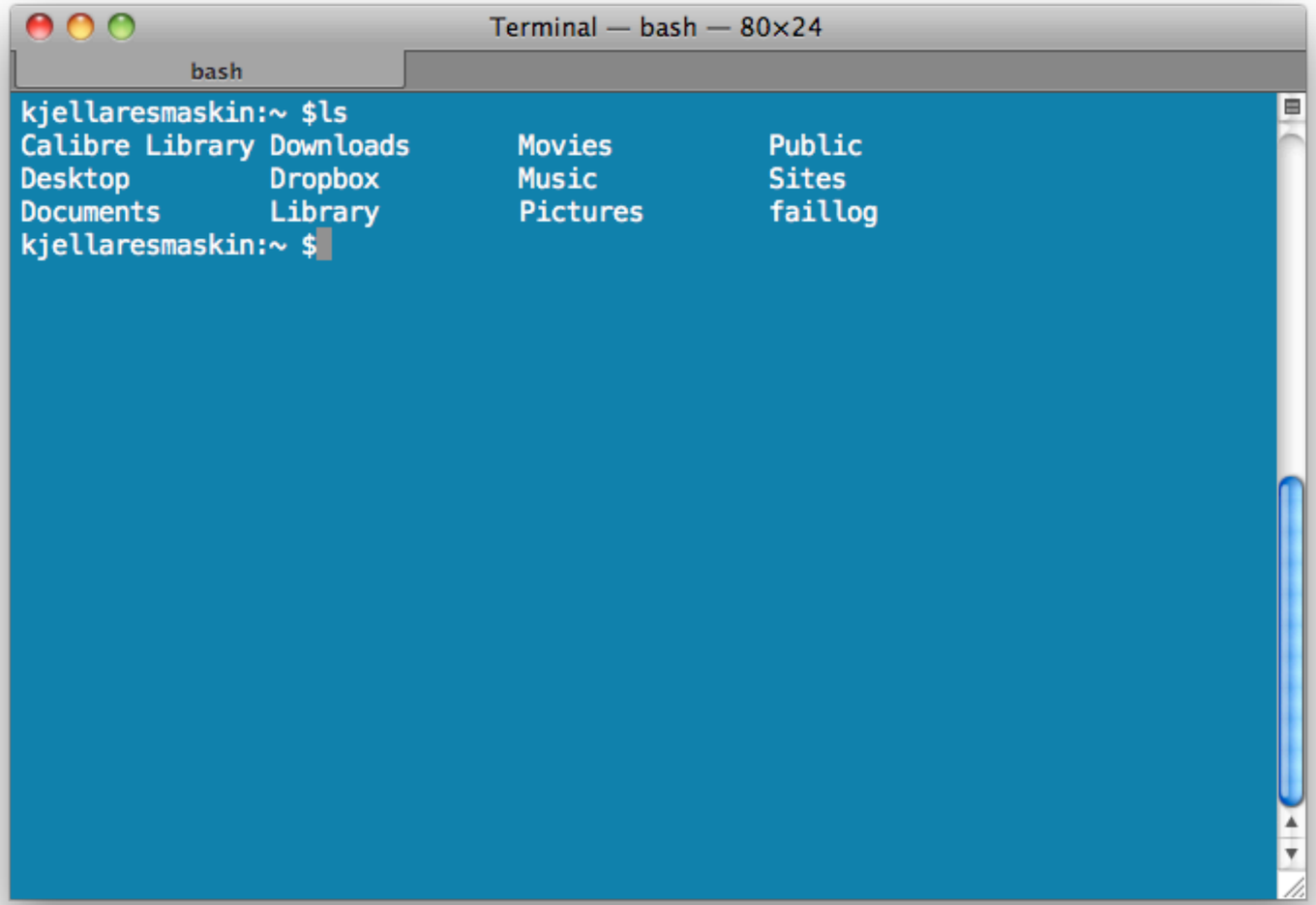
UNIX

- 1969, AT&T (USA/Bell Labs)
- Multitasking, multiuser, timesharing server/workstation operating system
- Popular in academic institutions (open, sharing) and later as a system that drove the machines on connected to the internet
- Unix philosophy: small simple programs that does one thing very well and can pipe the result to another program for further processing
- Still very popular and driving most new mobile devices
- UNIX have been re-created in the shape of GNU/Linux and there are a lot of versions/distributions built on the open/free Linux kernel today

UNIX

- UNIX™ owned by The Open Group (Technology Consortium)
- They certify systems to be branded as UNIX
- However, a lot of similar systems are branded “Unix-like” or “Un*x”
  - FreeBSD
  - Linux (Android)


Terminal — bash — 80×24
bash
kjellaresmaskin:~ $


Terminal — bash — 80×24
bash
kjellaresmaskin:~ $ls
Calibre Library Downloads Movies Public
Desktop Dropbox Music Sites
Documents Library Pictures faillog
kjellaresmaskin:~ $

ASSIGNMENTS

- Install a package management system on your Unix system to prepare for installing software
  - http://en.wikipedia.org/wiki/Package_management_system
- Begin by installing imagemagick, ffmpeg, latex/bibtex, gnuplot, exiftool. Some of these tools may not be available through a package management system but will come with their individual installers.
- Help eachother out using twitter (#imt4892) and meet up physically. Also - my office door is still open.

RESOURCES

Questions from these Wikipedia articles may end up in the final written exam.

- http://en.wikipedia.org/wiki/Computer_history
- http://en.wikipedia.org/wiki/History_of_computing_hardware
- http://en.wikipedia.org/wiki/Difference_engine
- http://en.wikipedia.org/wiki/History_of_the_graphical_user_interface
- http://en.wikipedia.org/wiki/Timeline_of_operating_systems
- http://en.wikipedia.org/wiki/Unix
- http://en.wikipedia.org/wiki/Graphical_user_interface
- http://en.wikipedia.org/wiki/Command_line_argument#Arguments
- http://en.wikipedia.org/wiki/Comparison_of_operating_systems
- http://en.wikipedia.org/wiki/Unix_philosophy
- http://en.wikipedia.org/wiki/List_of_Unix_utilities

CONTACT INFORMATION

Name	Kjell Are Refsvik
E-mail	kjell.refsvik@hig.no
Telephone	61 13 52 28 | 405 50 454
WWW	http://www.ansatt.hig.no/kjellr
Office	A224A-B
Available	https://hevn.hig.no/#user=kjell.refsvik@hig.no&view=cal

GJØVIK UNIVERSITY COLLEGE

Devices, formats, wrappers and compressors

Kjell Are Refsvik

September 7, 2011

IMT4892 Digital Workflow

http://www.ansatt.hig.no/kjellr/imt4892

20 WHERE WE ARE NOW

WEEK	TOPIC
33	-
34	Course Introduction
35	A Brief History of Computing and an Introduction to Unix
36	Devices and data formats, wrappers and compressors
37	Tools for transforming common datatypes
38	Moving data, remote processing and workflow automation
39	Workflow automation
40	Intellectual property rights (IPR), including Creative Commons
41	Mashup fundamentals and applied mashup using KML and Google Maps
42	XML Fundamentals, part 1
43	XML Fundamentals, part 2
44	Academic planning and writing
45	Project, week 1 - project outline delivery
46	Project, week 2 - work
47	Project delivery/presentation
48	Reading week before the exam
49	Exam

ASSIGNMENTS FROM LAST WEEK?

- Install a package management system on your Unix system to prepare for installing software
  - http://en.wikipedia.org/wiki/Package_management_system
- Begin by installing imagemagick, ffmpeg, latex/bibtex, gnuplot, exiftool. Some of these tools may not be available through a package management system but will come with their individual installers.
- Help eachother out using twitter (#imt4892) and meet up physically. Also - my office door is still open

TODAY'S GOAL

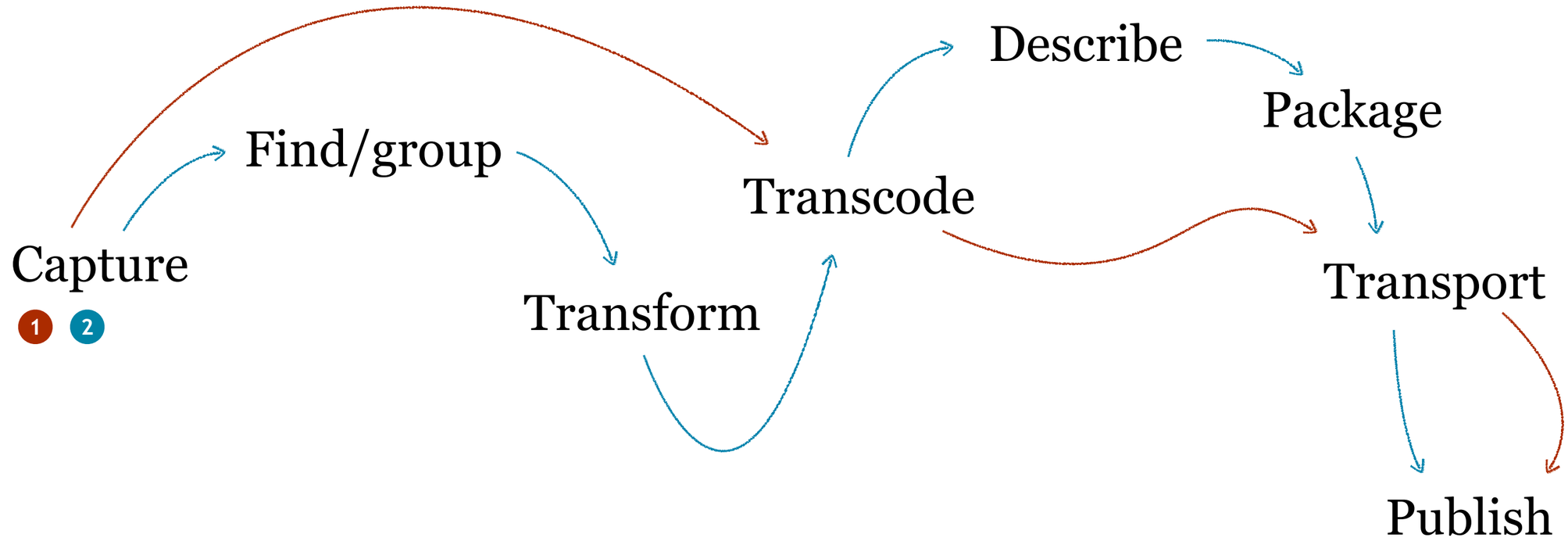
- After this week, you will know the difference between file compression method and a container format
- Know different groups of file formats and their attributes
- Know the importance of metadata in a workflow environment

CAPTURE DEVICES

CAPTURE DEVICES

- Capture and store data
- Typical ones are devices that capture images, sound, geo (gps)
- Format can be raw or processed
- Processed (jpg/mp3) can be compressed with a lossy compression algoritm and processing non-reversable. Beware of quality loss.
- Some modern devices today can be set up as a complete workflow in and of itself. Cameras/smartphones.

Can sometimes be set up to be a complete workflow.



CAPTURE DEVICES

- When choosing one: look for a devicer that can capture high quality:
  - Data (i.e. images)
  - Metadata (data about the data). Location, date, time, face recog., compass, location.
- Metadata important later on as you want to search for images with specific characteristics.

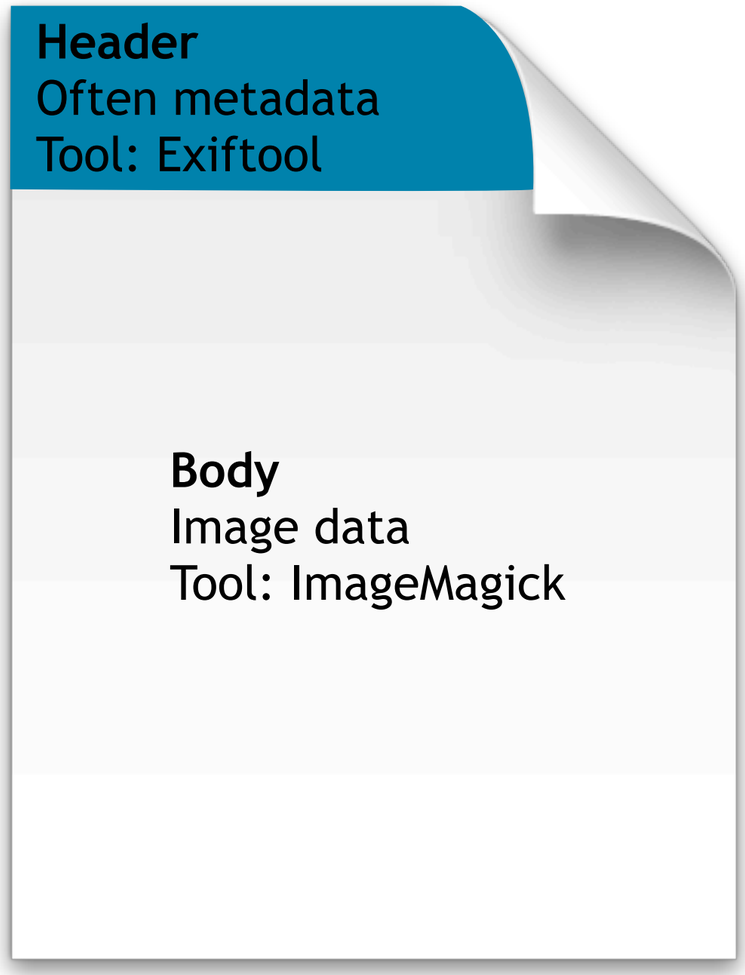
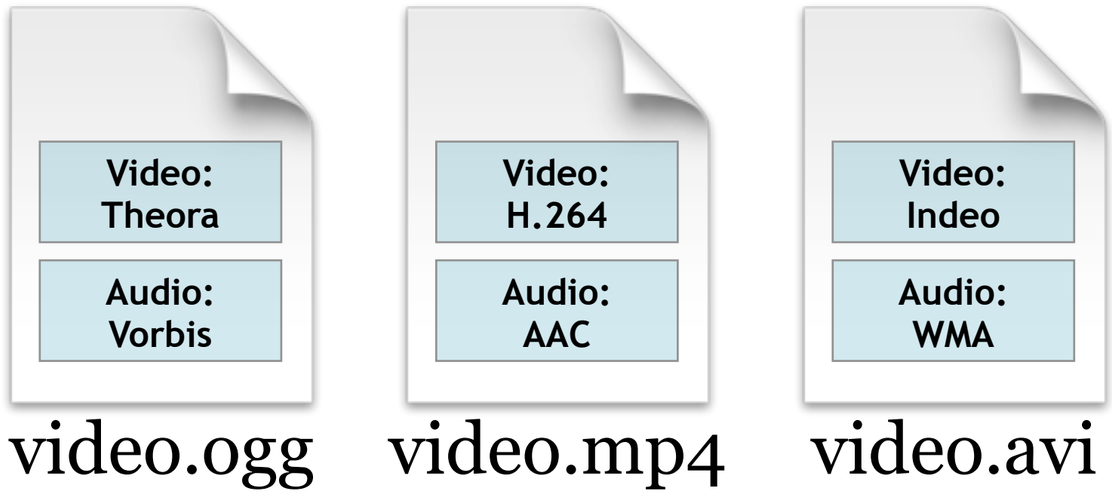

Header
Often metadata
Tool: Exiftool
Body
Image data
Tool: ImageMagick

image.jpg

WRAPPER FORMATS

WRAPPER FORMATS

- Purpose: wrapping more than a single data format
- Typical example: video (sound+images)
- Wrapper format examples: OGG, MP4, AVI
- Inside them, compressed audio and video.



COMPRESSION ALGORITMS

COMPRESSION ALGORITMS

- Purpose: compress data more to a more managable size that is can be moved and stored faster and cheaper
- Two main categories: lossless and lossy
  - Lossless: compresses data without loosing quality
  - Lossy: compresses data by removing data

bob.cr2
raw
3888x2592px
13,8MB

bob.jpg
sRGB
3888x2592px
2,5MB

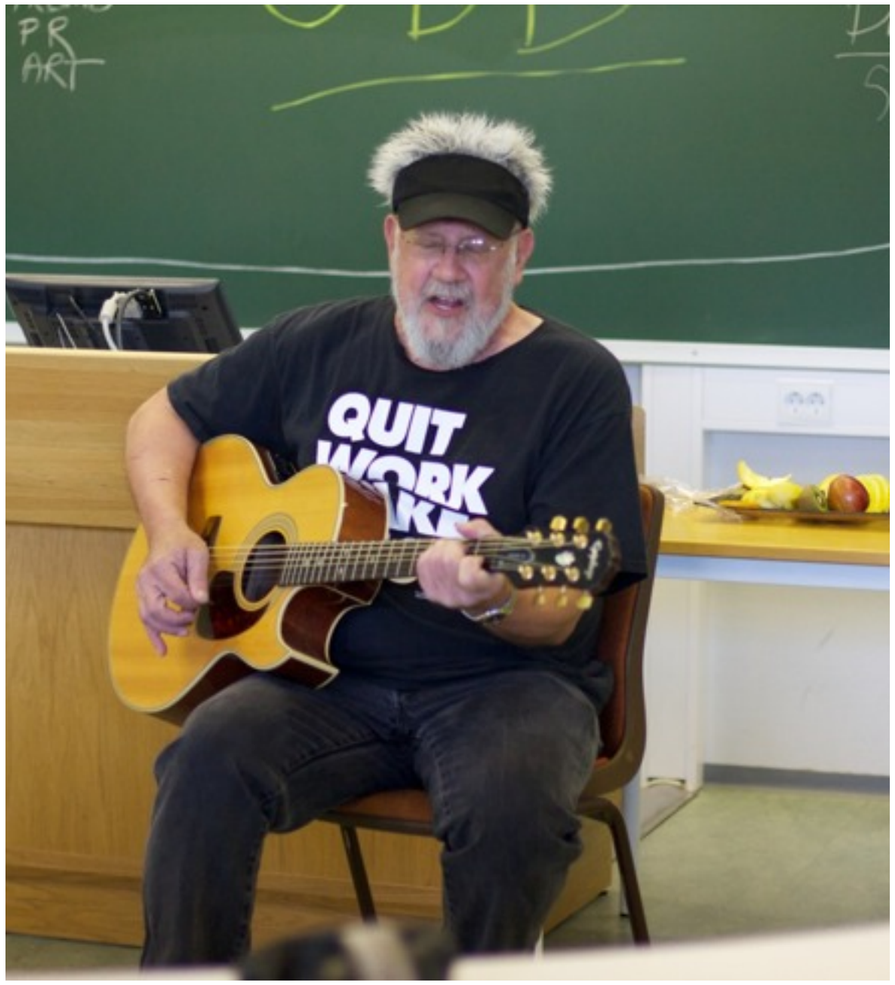
bob.tiff

Lossy compression

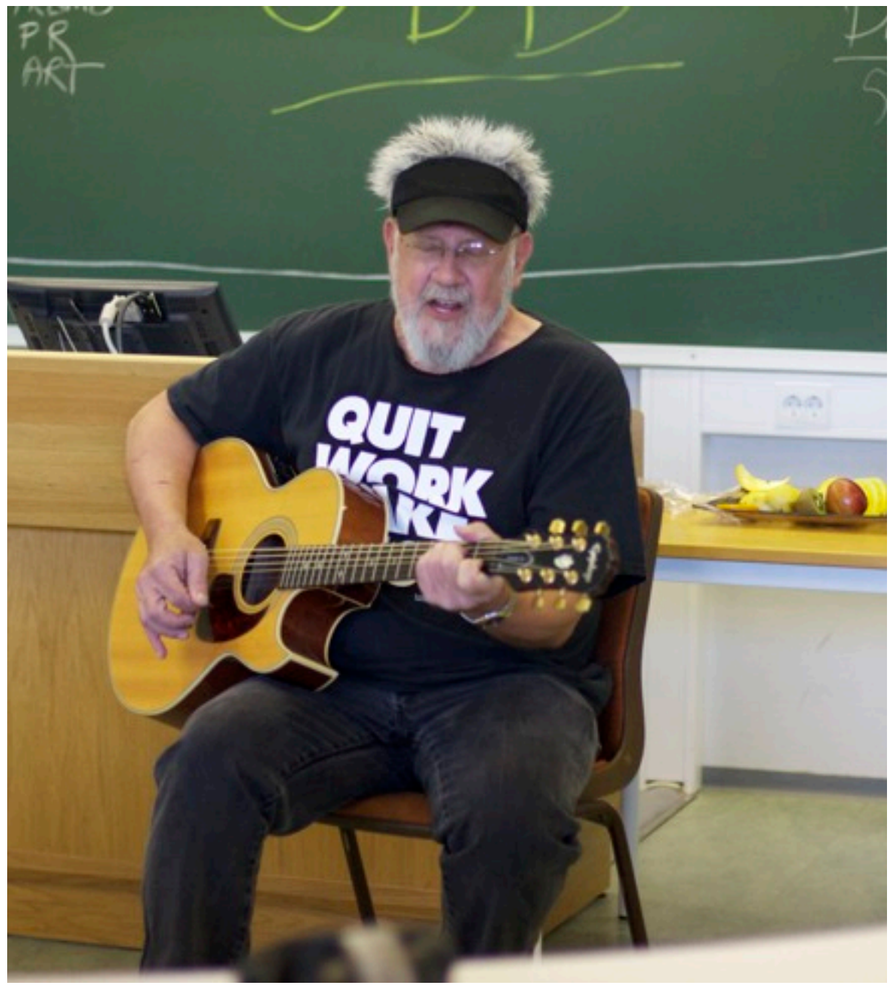
bob.png
sRGB
3888x2592px
16MB

bob.tiff

Lossless compression



bob.jpg | sRGB | 3888x2592px | 2,5MB

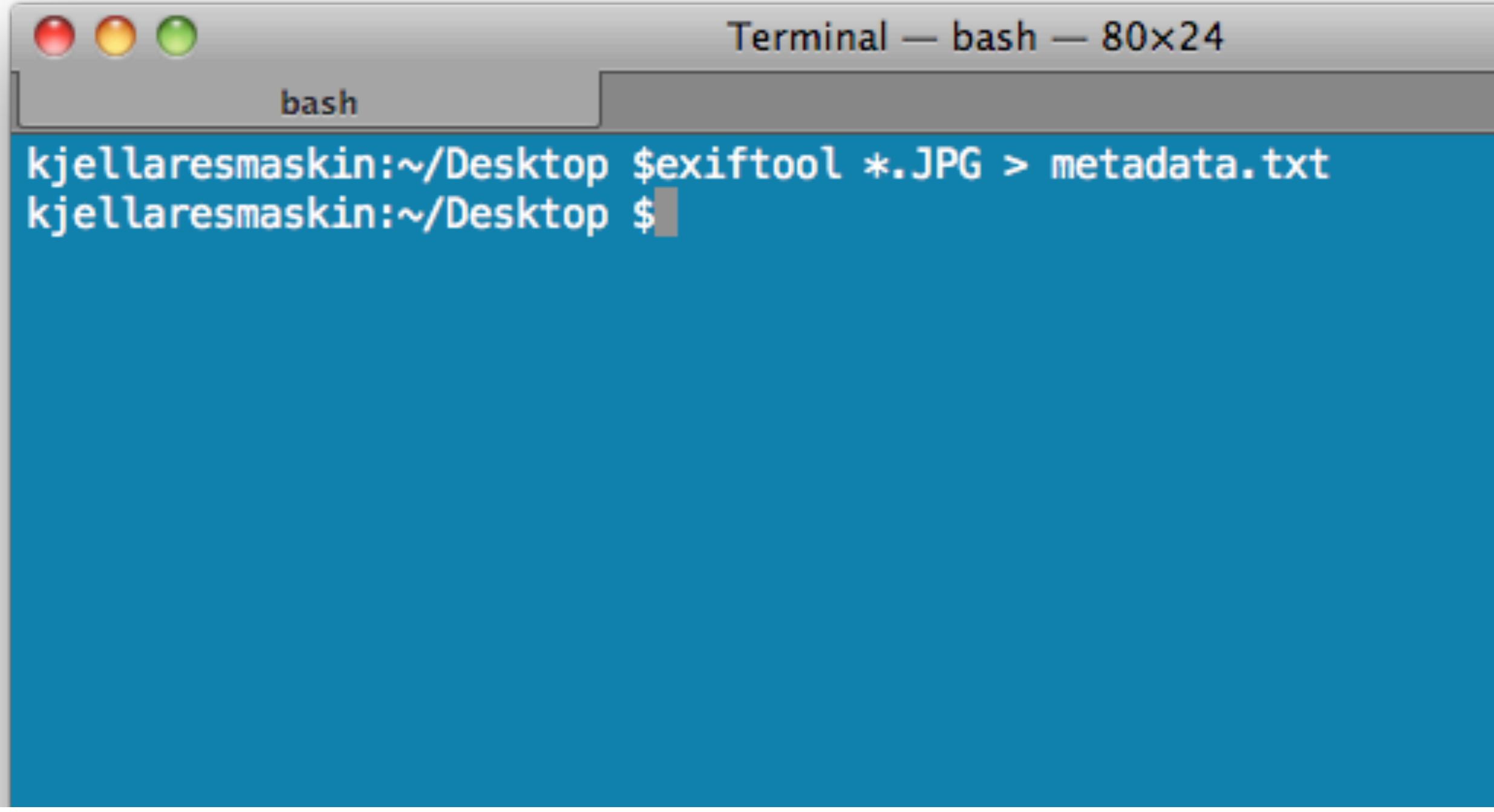


bob.cr2 | raw | 3888x2592px | 13,8MB

METADATA

METADATA

- Data describing data
- Often found in the file-header
- Filenames can also carry metadata
- Should be on your watch-list when buying devices
- Collecting a lot of metadata could you organise and find files and important when file numbers grow


Terminal — bash — 80×24
bash
kjellaresmaskin:~/Desktop $exiftool *.JPG > metadata.txt
kjellaresmaskin:~/Desktop $

```
ExifTool Version Number         : 7.93
File Name                       : IMG_0338.JPG
Directory                       : .
File Size                       : 2.2 MB
File Modification Date/Time     : 2010:08:31 14:52:56+02:00
File Type                       : JPEG
MIME Type                       : image/jpeg
Exif Byte Order                 : Big-endian (Motorola, MM)
Image Description               : Back Camera
Make                            : Apple
Camera Model Name               : iPhone
Orientation                     : Horizontal (normal)
X Resolution                    : 72
Y Resolution                    : 72
Resolution Unit                 : inches
Software                        : 4.0.2
Modify Date                     : 2010:08:31 14:52:56
Y Cb Cr Positioning             : Centered
Exposure Time                   : 1/1436
F Number                        : 2.4
Exposure Program                : Program AE
```

```
ISO                                 : 80
Exif Version                        : 0221
Date/Time Original                  : 2010:08:31 14:52:56
Create Date                         : 2010:08:31 14:52:56
Components Configuration            : Y, Cb, Cr, -
Shutter Speed Value                 : 1/1436
Aperture Value                      : 2.4
Metering Mode                       : Average
Flash                               : Auto, Did not fire
Focal Length                        : 3.9 mm
Subject Area                        : 1295 967 699 696
Flashpix Version                    : 0100
Color Space                         : sRGB
Exif Image Width                    : 0
Exif Image Height                   : 0
Sensing Method                      : One-chip color area
Exposure Mode                       : Auto
White Balance                       : Auto
Scene Capture Type                  : Standard
Sharpness                           : Hard
```

```
GPS Latitude Ref                : North
GPS Longitude Ref               : East
GPS Time Stamp                  : 14:52:50.78
GPS Img Direction Ref           : True North
GPS Img Direction               : 287.5587302
Compression                     : JPEG (old-style)
Thumbnail Offset                : 884
Thumbnail Length                : 9701
Image Width                     : 2592
Image Height                    : 1936
Encoding Process                : Baseline DCT, Huffman coding
Bits Per Sample                 : 8
Color Components                : 3
Y Cb Cr Sub Sampling            : YCbCr4:2:0 (2 2)
Aperture                        : 2.4
GPS Latitude                    : 60 deg 47' 25.80" N
GPS Longitude                   : 10 deg 41' 0.00" E
GPS Position                    : 60 deg 47' 25.80" N, 10 deg 41' 0.00" E
Image Size                      : 2592x1936
Shutter Speed                   : 1/1436
```

```
Thumbnail Image                 : (Binary data 9701 bytes, use -b option
to extract)
Focal Length                    : 3.9 mm
Light Value                     : 13.3
```

IDENTIFICATION OF FILES

- Internet Assigned Numbers Authority
  - http://www.iana.org/assignments/media-types/
- Internet Engineering Task Force
  - http://en.wikipedia.org/wiki/Internet_media_type
  - http://en.wikipedia.org/wiki/Filename_extension
  - http://www.file-extensions.org/

FILE FORMATS

- Closed, often binary and tied to a specific vendor to serve the vendor´s needs (be careful of vendor lock-in situations and dependencies). Examples: .doc, .xls, .ppt, .fla, .cr2, .nef
- Open, often standardized by ISO, W3C, IEEE or other standization organisations to be partly or completely open to everyone to use. Examples: svg, html, kml, css, jpg, tiff

Flash vs HTML5/CSS3/Javascript?

Sustainability

“Digital sustainability is the collection, processing, and presentation of our data today - done without limiting future generations to access and manipulate the same data.”

Freely adapted from a definition of “sustainability” in general, used in the Bruntland report, “Our common Future”, 1987
http://en.wikipedia.org/wiki/Our_Common_Future

Sustainability factors - file formats

Library of Congress (USA)
http://www.digitalpreservation.gov/formats/

Disclosure
Adoption
Transparency
Self-documentation
External dependencies
Impact of patents
Technical protection mechanisms

Sustainability factors - file formats

Norwegin Ministry of Government Administration, Reform and Church Affairs:
http://www.regjeringen.no/en/dep/fad/press-centre/press-releases/2009/new-obligatory-it-standards-for-the-stat.html?id=570650

In short, this is the content of the government decision:

- On government operated web sites, from 1.1.2012 it will be obligatory to publish multimedia content in open formats:
  - For video: Theora/Vorbis/Ogg or H.264/AAC/MP4.
  - For sound: Vorbis/Ogg, MP3 or FLAC/Ogg.
  - For pictures: JPEG or PNG.
- When exchanging documents as attachments to e-mail between government institutions and users, from 1.1.2011 it will be obligatory to use the document formats PDF or ODF.
- Version change: From 1.1.2010 the ODF version 1.1 is to be used.
- The standard for character sets ISO10646, represented by UTF8, is to be used at all new ICT projects in the government sector. From 1.1.2012, UTF8 is to be used during electronic information exchange. It will possible to make exemptions from this demand in special cases.

The UTF8 decision is an important step in enabling the public sector to handle characters in the Sami language and other languages in a correct way. At first, the obligatory demands will only be applied to the state sector. The Ministry of Government Administration and Reform is planning regulations to make the obligatory standards applicable for the municipal sector as well.

ASSIGNMENTS

- Connect a prefered device to your UNIX machine (i.e. camera) and download data from the device to your machine
- Transcode the data into different data format (i.e. JPEG to PNG) using imagemagick (see documentation online)
- Read through wiki articles in this lecture

RESOURCES

- http://en.wikipedia.org/wiki/Data_compression
- http://en.wikipedia.org/wiki/Comparison_of_container_formats
- http://en.wikipedia.org/wiki/Container_format_(digital)
- http://en.wikipedia.org/wiki/File_format
- http://en.wikipedia.org/wiki/Vendor_lock-in
- http://en.wikipedia.org/wiki/Data_compression#Compression_algorithms

CONTACT INFORMATION

Name	Kjell Are Refsvik
E-mail	kjell.refsvik@hig.no
Telephone	61 13 52 28 | 405 50 454
WWW	http://www.ansatt.hig.no/kjellr
Office	A224A-B
Available	https://hevn.hig.no/#user=kjell.refsvik@hig.no&view=cal

GJØVIK UNIVERSITY COLLEGE

Tools for transforming common datatypes

Kjell Are Refsvik

September 14, 2011

IMT4892 Digital Workflow

http://www.ansatt.hig.no/kjellr/imt4892

20 WHERE WE ARE NOW

WEEK	TOPIC
33	-
34	Course Introduction
35	A Brief History of Computing and an Introduction to Unix
36	Devices and data formats, wrappers and compressors
37	Tools for transforming common datatypes
38	Moving data, remote processing and workflow automation
39	Workflow automation
40	Intellectual property rights (IPR), including Creative Commons
41	Mashup fundamentals and applied mashup using KML and Google Maps
42	XML Fundamentals, part 1
43	XML Fundamentals, part 2
44	Academic planning and writing
45	Project, week 1 - project outline delivery
46	Project, week 2 - work
47	Project delivery/presentation
48	-
49	Exam

ASSIGNMENTS FROM LAST WEEK?

- Connect a prefered device to your UNIX machine (i.e. camera) and download data from the device to your machine
- Transcode the data into different data format (i.e. JPEG to PNG) using imagemagick (see documentation online)
- Read through wiki articles in this lecture

TODAY'S GOAL

- Know how the difference between transforming and transcoding data
- Learn how to transform and transcode common file-based datatypes

INFORMATION

TO NEW STUDENTS

The Reqruitment Survey for 2011 are now sent out to all first semester students at the bachelor level and students following one-year programmes.

Topics are demography, information retrieval, choice of study, the admission process, semester start, the first weeks at GUC and motivation, among other things.

Sincerely,
Gro Jeppestøl Vatne
Department of Academic Affairs
Gjøvik University College

CODE

A code is a rule for converting a piece of information (for example, a letter, word, phrase, or gesture) into another form or representation (one sign into another sign), not necessarily of the same type.

Source: http://en.wikipedia.org/wiki/Code

ENCODING/DECODING

In communications and information processing,encoding is the process by which information from a source is converted into symbols to be communicated.

Decoding is the reverse process, converting these code symbols back into information understandable by a receiver.

Source: http://en.wikipedia.org/wiki/Encoder

TRANSCODE

Transcoding is the direct digital-to-digital data conversion of one encoding to another,[1] such as for movie data files or audio files.

This is usually done in cases where a target device (or workflow) does not support the format or has limited storage capacity that mandates a reduced file size,[1] or to convert incompatible or obsolete data to a better-supported or modern format.

Source: http://en.wikipedia.org/wiki/Transcoding

TRANSFORM

Transform coding is a type of data compression for "natural" data like audio signals or photographic images. The transformation is typically lossy, resulting in a lower quality copy of the original input.

In transform coding, knowledge of the application is used to choose information to discard, thereby lowering its bandwidth. The remaining information can then be compressed via a variety of methods. When the output is decoded, the result may not be identical to the original input, but is expected to be close enough for the purpose of the application.

Source: http://en.wikipedia.org/wiki/Transform_coding

TRANSCODING

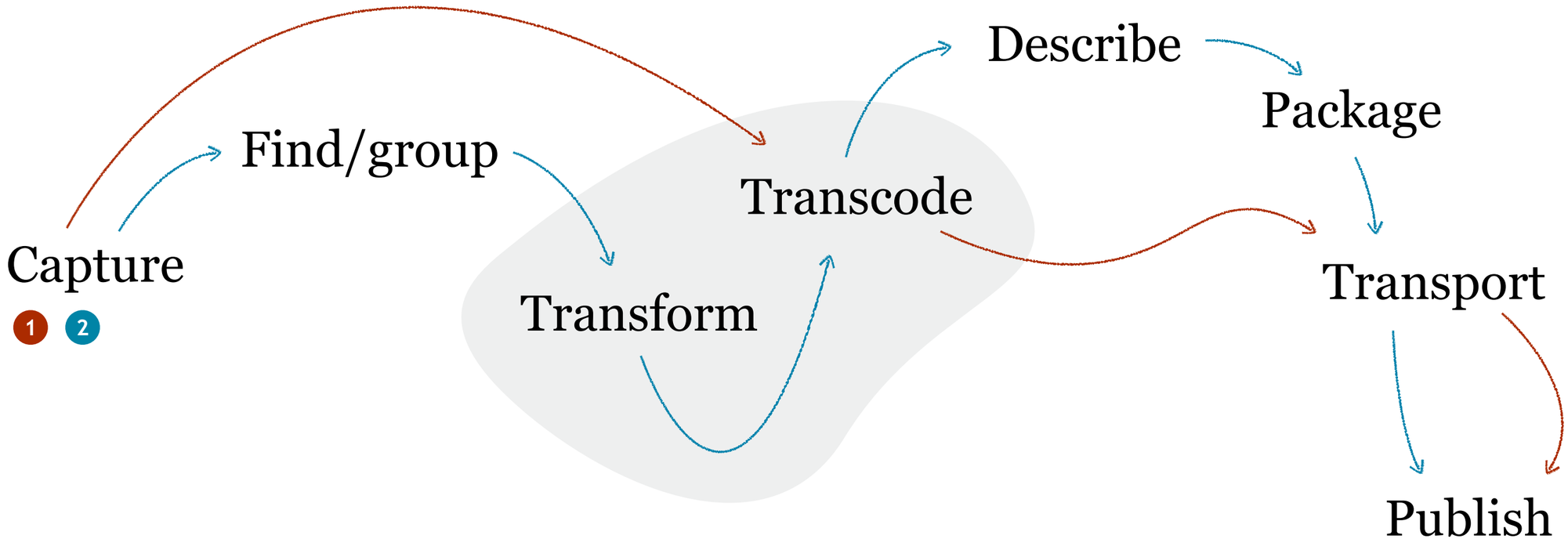
Transcoding is the direct digital-to-digital data conversion of one encoding to another,[1] such as for movie data files or audio files.

This is usually done in cases where a target device (or workflow) does not support the format or has limited storage capacity that mandates a reduced file size,[1] or to convert incompatible or obsolete data to a better-supported or modern format.

Source: http://en.wikipedia.org/wiki/Transcoding

SUMMATION

- Transcoding: JPG > PNG
- Transforming: 3000x2000 px > 640x480 px
- A digital workflow often includes both
- The tools we have selected for media processing can do both


Capture
1
2
Find/group
Transform
Transcode
Describe
Package
Transport
Publish

Toolbox

Text	Images/graphs	Geodata	Audio	Video	Metadata
grep	imagemagick	gpsbabel	ffmpeg	ffmpeg	exiftool
sed	gnuplot			mplayer	
awk	dcraw			vlc	
cut					
find					
cat					

For more software - go to your package installer repository to look for more

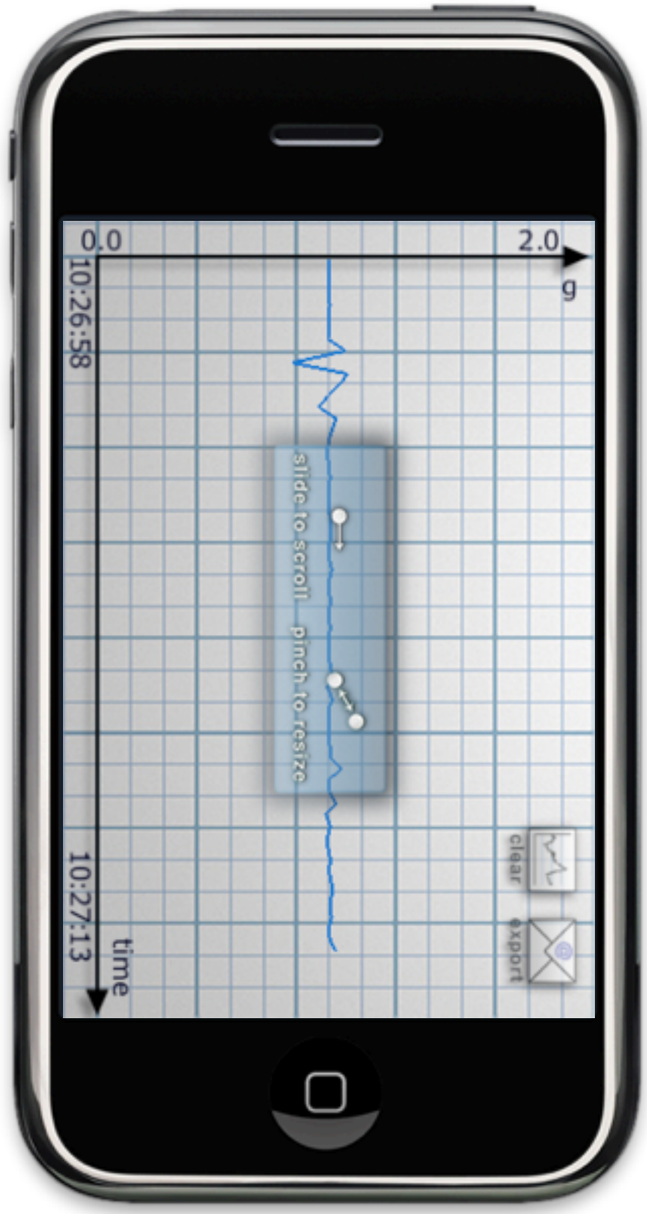
CHARACTER ENCODING IN TEXT

- System of bitstreams to describe alfa-numeric (and other) characters that you type. Many systems out there.
  - http://en.wikipedia.org/wiki/Character_encoding
- Please use Unicode at ALL times (UTF-8)
  - http://en.wikipedia.org/wiki/UTF-8
- Web pages:
  - **store** html/css files as UTF-8 encoded text
  - **declare** UTF-8 and character encoding method in the top of HTML doc.
  - make your web server **deliver** UTF-8 characters

Text


340 350 N 10 20
SPEED
KM/H
ALTITUDE
METER
ACCELERATION
g UNITS
10:27:08
- m
- km/h
60°47'24"N
10°40'60"E
GPS SIGNAL
MASTER ALERT
PULL UP
MUTE
NIGHT VISION
CAL
SETUP
HELP


ACCELERATION
g UNITS


0.0
2.0
g
10:26:58
slide to scroll
pinch to resize
clear
export
time
10:27:13

g_force.csv

```
Time;Speed (km/h)
07:31:22;0.150169
07:31:23;0.075085
07:31:24;0.037542
07:31:25;0.018771
07:31:26;0.009386
07:31:27;0.004693
07:31:28;0.002346
07:31:29;0.001173
07:31:30;0.000587
07:31:31;0.000293
07:31:32;0.000147
07:31:33;0.000073
07:31:34;0.000037
07:31:35;0.000018
```

g_force.csv

```
Time;Speed (km/h)
07:31:22;0.150169
07:31:23;0.075085
07:31:24;0.037542
07:31:25;0.018771
07:31:26;0.009386
07:31:27;0.004693
07:31:28;0.002346
07:31:29;0.001173
07:31:30;0.000587
07:31:31;0.000293
07:31:32;0.000147
07:31:33;0.000073
07:31:34;0.000037
07:31:35;0.000018
```

Terminal — less — 92×25

less

```
kjellaresmaskin:~/Desktop $cut -f 2 -d ';' g_force.csv | sed '1d' > g_force_processed.csv
kjellaresmaskin:~/Desktop $
```

```
Time;Speed (km/h)          0.150169
07:31:22;0.150169          0.075085
07:31:23;0.075085          0.037542
07:31:24;0.037542          0.018771
07:31:25;0.018771          0.009386
07:31:26;0.009386          0.004693
07:31:27;0.004693          0.002346
07:31:28;0.002346          0.001173
07:31:29;0.001173          0.000587
07:31:30;0.000587          0.000293
07:31:31;0.000293          0.000147
```

Terminal — bash — 92×25

bash

```
kjellaresmaskin:~/Desktop $cut -f 2 -d ';' g_force.csv | sed '1d' > g_force_processed.csv
kjellaresmaskin:~/Desktop $gnuplot

        G N U P L O T
        Version 4.4 patchlevel 0
        last modified March 2010
        System: Darwin 10.4.0

        Copyright (C) 1986-1993, 1998, 2004, 2007-2010
        Thomas Williams, Colin Kelley and many others

        gnuplot home:     http://www.gnuplot.info
        faq, bugs, etc:   type "help seeking-assistance"
        immediate help:   type "help"
        plot window:      hit 'h'

Terminal type set to 'aqua'
gnuplot> plot 'g_force_processed.csv' with lines
gnuplot>
```

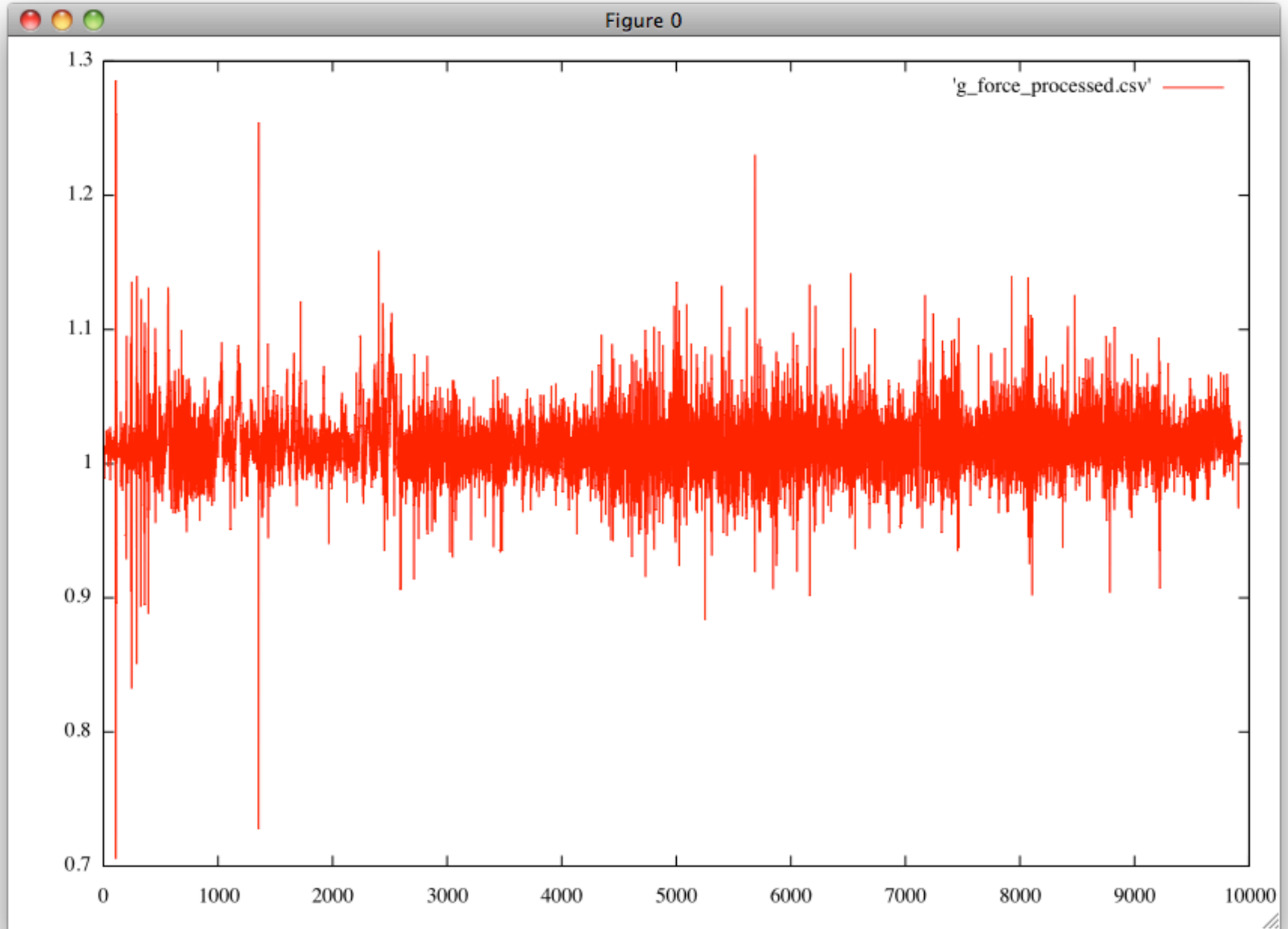


Figure 0
'g_force_processed.csv'
1.3
1.2
1.1
1
0.9
0.8
0.7
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000

Text manipulation

- Important to know
- A whole lot of basic text-manipulation tools in UNIX:
  sed, ed, grep, awk, find, cut, sort, tr,
- Read up on them in the book
- Create a text-file with useful command-lines to that you can remember them and easily create new ones

Images

ImageMagick - very powerful tool for transforming and/or transcoding image data

```
mogrify -resize 50% -format jpg *.PNG
```

[transmogrify]

destination source files

http://www.imagemagick.org/script/command-line-tools.php
for more information on imagemagick commands and parameters.

Metadata

ExifTool - very powerful tool for transforming and/or transcoding metadata

```
exiftool "-Artist=Kjell Are Refsvik" -overwrite_original *.jpg
```

http://www.sno.phy.queensu.ca/~phil/exiftool/#running
for more information on ExifTool commands and parameters.

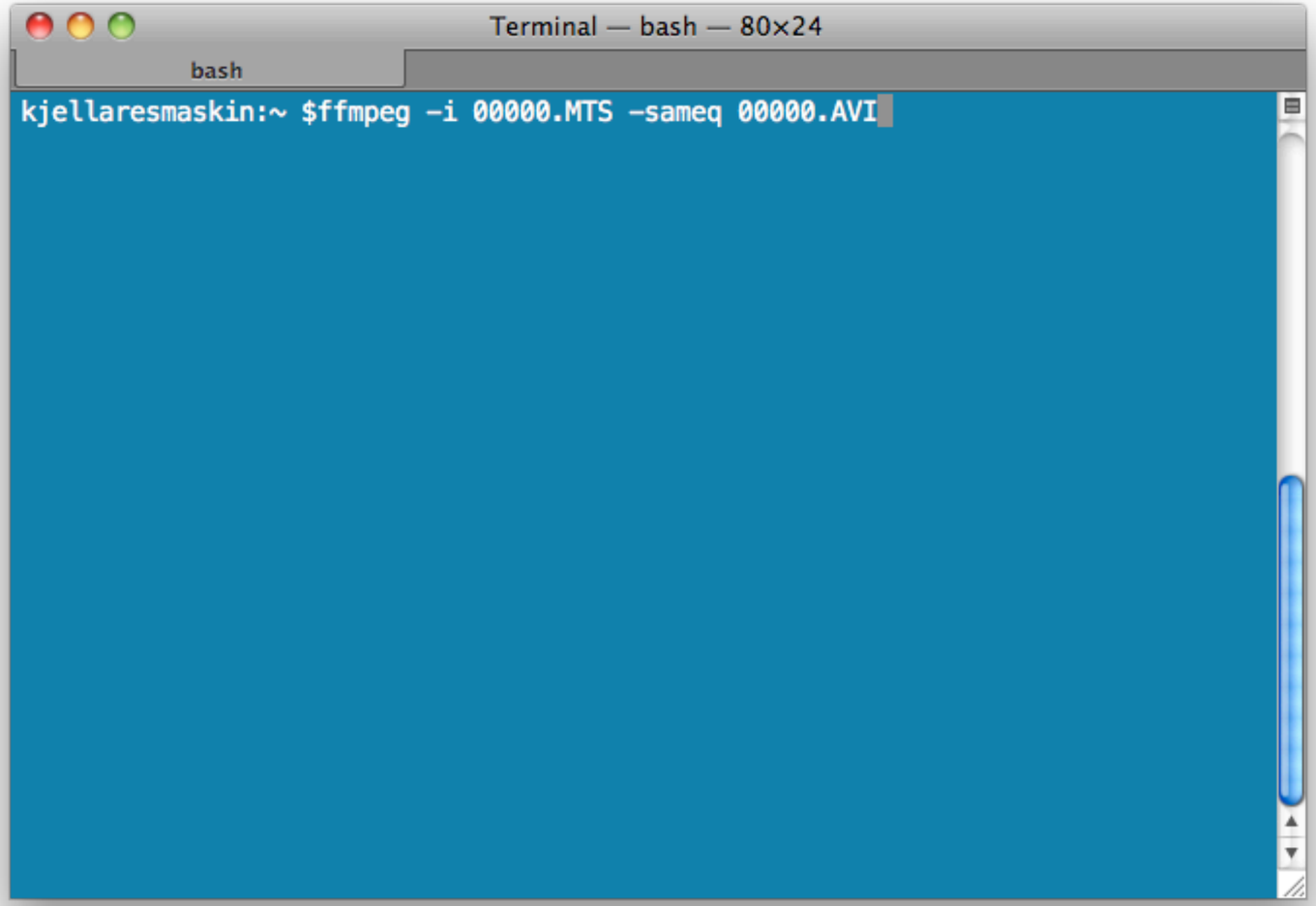

EGON
RESTAURANT
GJØVIK
Håndballklubb
SUBARU
SAAB
ØYHUS MOTOR AS
Oppland
ARBEIDERBLAD
radiometro
radiometro
Coca-Cola
Sparebanken 1
HEDMARK

```
ExifTool Version Number         : 8.29
File Name                       : 00000.MTS
Directory                       : .
File Size                       : 15 MB
File Modification Date/Time     : 2009:01:30 21:02:20+01:00
File Permissions                : rwxrwxrwx
File Type                       : M2TS
MIME Type                       : video/m2ts
Video Stream Type               : H.264 Video
Audio Stream Type               : A52/AC-3 Audio
Audio Bitrate                   : 256000
Surround Mode                   : Not indicated
Audio Channels                  : 2
Image Width                     : 1440
Image Height                    : 1080
Date/Time Original              : 2009:01:30 22:02:01+00:00
Aperture Setting                : Auto
Gain                            : 3 dB
Exposure Program                : Program AE
White Balance                   : Auto
Focus                           : Auto (0.045)
Image Stabilization             : On
F Number                        : 1.8
Make                            : Sony
Audio Sample Rate               : 48000
Aperture                        : 1.8
Image Size                      : 1440x1080
```

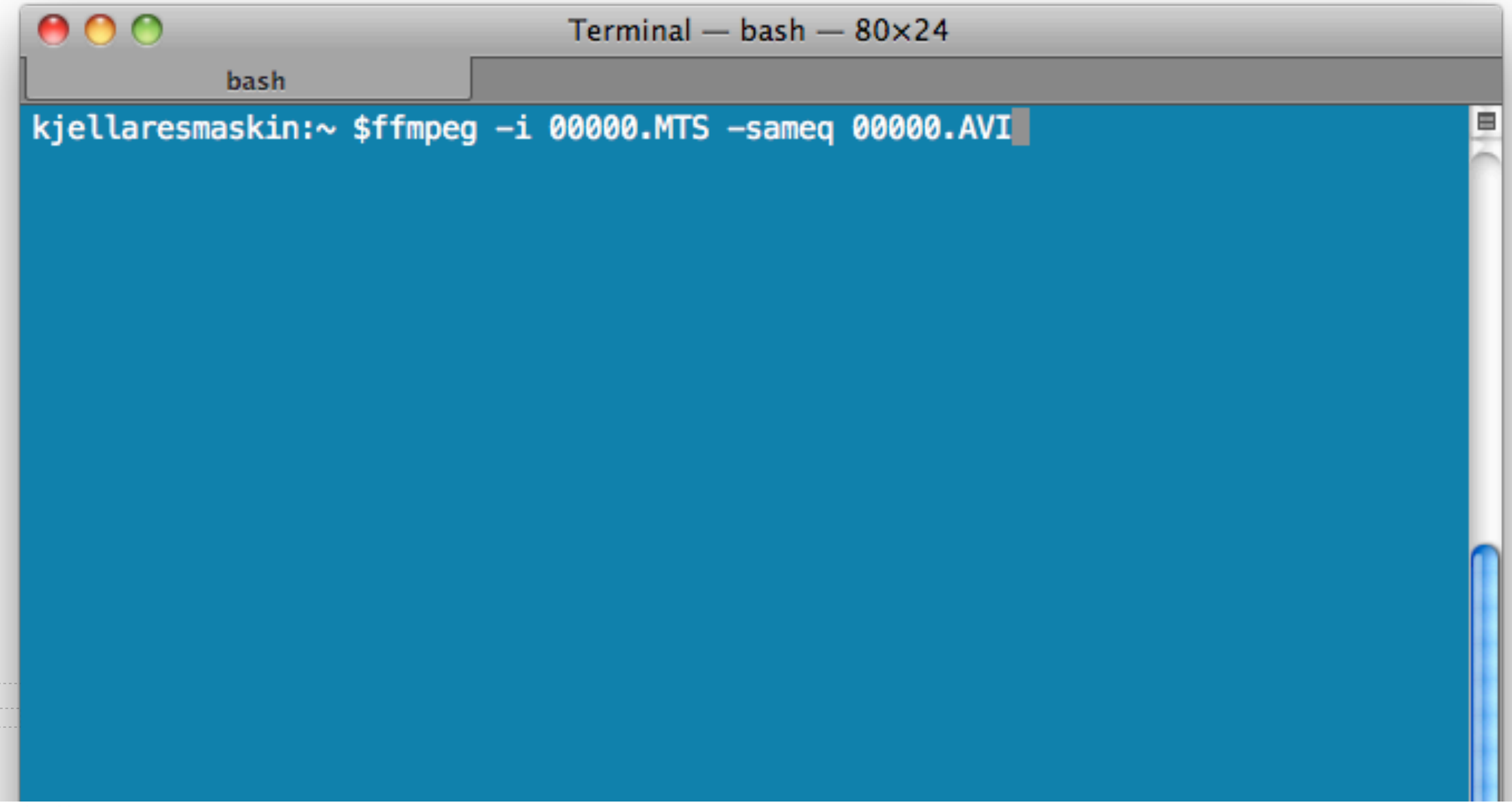
00000.MTS **00000.AVI**

```
ExifTool Version Number         : 8.29
File Name                       : 00000.AVI
Directory                       : .
File Size                       : 49 MB
File Modification Date/Time     : 2011:09:12 13:55:49+02
File Permissions                : rw-r--r--
File Type                       : AVI
MIME Type                       : video/avi
Frame Rate                      : 50
Max Data Rate                   : 32.23 kB/s
Frame Count                     : 937
Stream Count                    : 2
Stream Type                     : Video
Video Codec                     : FMP4
Video Frame Rate                : 50
Video Frame Count               : 937
Quality                         : Default
Sample Size                     : Variable
Image Width                     : 1440
Image Height                    : 1080
Planes                          : 1
Bit Depth                       : 24
Compression                     : FMP4
Image Length                    : 4665600
Pixels Per Meter X              : 0
Pixels Per Meter Y              : 0
Num Colors                      : Use BitDepth
Num Important Colors            : All
Audio Codec                     : .
Audio Sample Rate               : 41.67
Audio Sample Count              : 780
Encoding                        : Microsoft MPEG
Num Channels                    : 2
Sample Rate                     : 48000
Avg Bytes Per Sec               : 8000
Bits Per Sample                 : 0
Software                        : Lavf52.64.2
Duration                        : 18.74 s
Image Size                      : 1440x1080
```

VIDEO


Terminal — bash — 80×24
bash
kjellaresmaskin:~ $ffmpeg -i 00000.MTS -sameq 00000.AVI

VIDEO


Terminal — bash — 80×24
bash
kjellaresmaskin:~ $ffmpeg -i 00000.MTS -sameq 00000.AVI

Video

7+ hrs television programme about the trainride from Bergen to Oslo (500+ km).

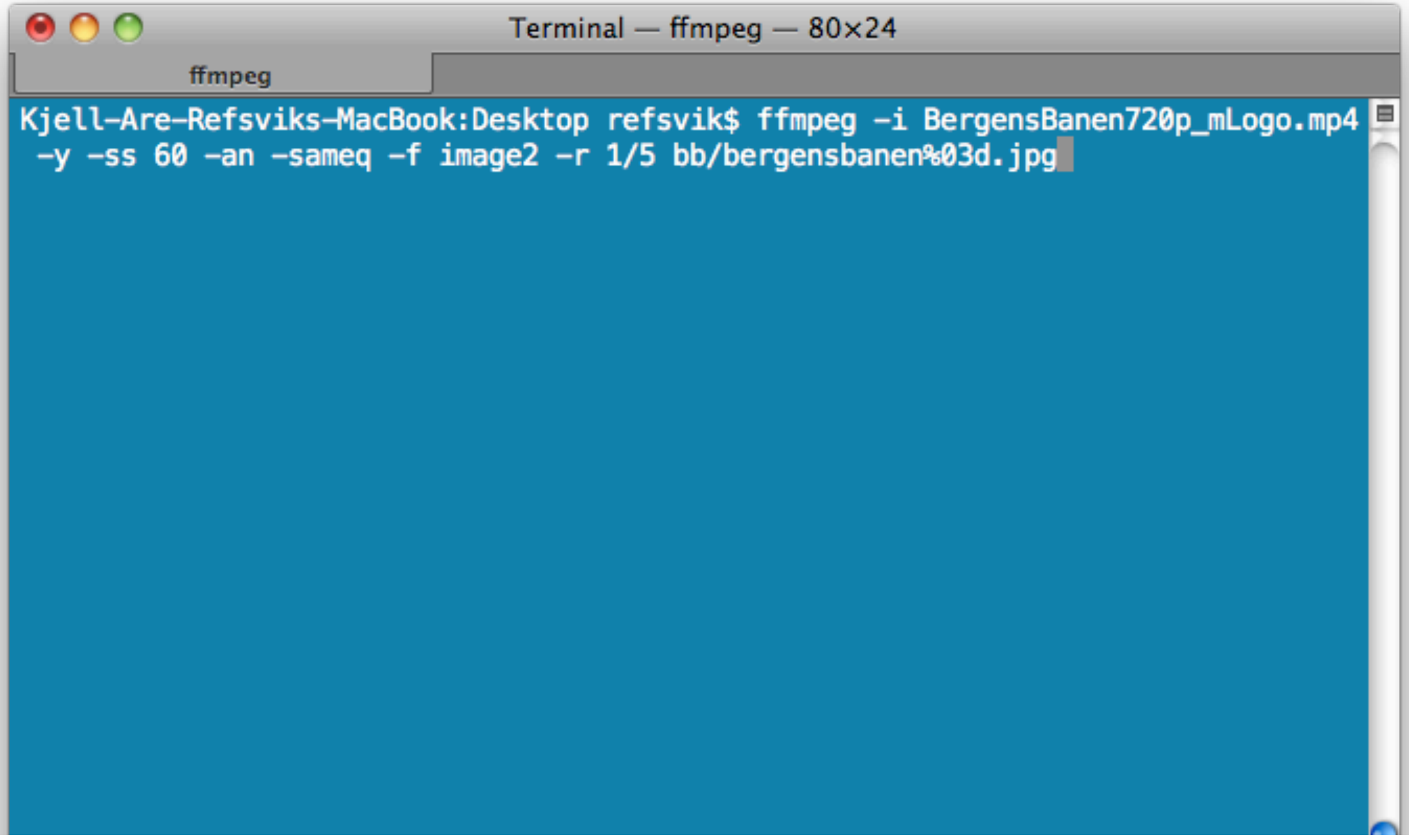
22GB file shared on the web.

How do we make a short/quick version using UNIX tools?

Destination:

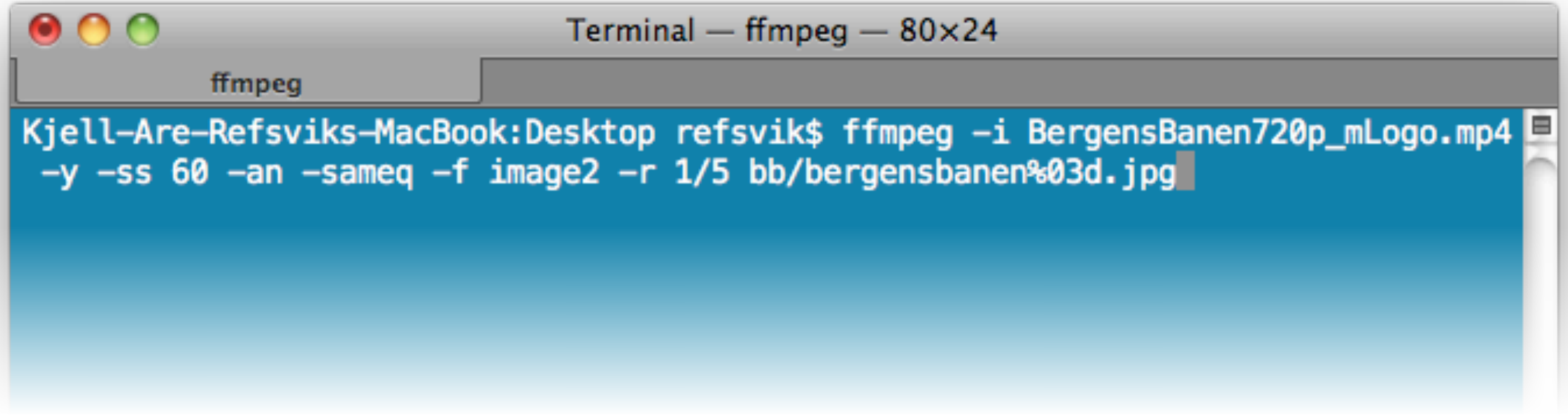
http://www.ansatt.hig.no/kjellr/misc/20100128_Bergensbanen_0328.html

Source: http://nrkbeta.no/2009/12/18/bergensbanen-eng/


Terminal — ffmpeg — 80×24
ffmpeg
Kjell-Are-Refsviks-MacBook:Desktop refsvik$ ffmpeg -i BergensBanen720p_mLogo.mp4
 -y -ss 60 -an -sameq -f image2 -r 1/5 bb/bergensbanen%03d.jpg

```
ffmpeg -i BergensBanen720p_mLogo.mp4 -y -ss 60 -an -sameq -f image2 -r 1/5 bb/bergensbanen%03d.jpg
```

the program we are running	input filename	over-write output files	seek 60 seconds out for the next frame	audio = null	dest. quality same as source	output format	frame-rate = 1/5 frame per second	output filename with option to add a frame number for each jpg file.



A lot of potential here to make this smarter, and also have it create the final .mp4 when it is done.

Audio

ffmpeg - very powerful tool for transforming and/or transcoding audio data

```
ffmpeg -b 192k -i file.wav file.mp3
```

http://ffmpeg.org/
for more information on gpsbabel commands and parameters.

Geodata

gpsbabel - very powerful tool for transforming and/or transcoding gps data

- `gpsbabel -t -i gpx -f trip.gpx -x track,start=2010090709,stop=2010090719 -o kml -F out.kml`

http://gpsbabel.org
for more information on gpsbabel commands and parameters.

Geodata

gpsbabel - very powerful tool for transforming and/or transcoding gps data

- `gpsbabel -t -i gpx -f trip.gpx -x track,start=2010090709,stop=2010090719 -o kml -F out.kml`

software

read track data

input format is .gpx

input filename

extract data between start and stop

output format is .kml

output filename is “out.kml”

http://gpsbabel.org
for more information on gpsbabel commands and parameters.

ASSIGNMENTS

- Try to transform and transcode different data-types (images, geo, video, audio)

RESOURCES

- http://www.gpsbabel.org/readme.html
- http://www.imagemagick.org/script/command-line-tools.php?ImageMagick=vg5hihs3gg8m5gh684ggul07i4
- http://ffmpeg.org/documentation.html
- http://www.sno.phy.queensu.ca/~phil/exiftool/

CONTACT INFORMATION

Name	Kjell Are Refsvik
E-mail	kjell.refsvik@hig.no
Telephone	61 13 52 28 | 405 50 454
WWW	http://www.ansatt.hig.no/kjellr
Office	A224A-B
Available	https://hevn.hig.no/#user=kjell.refsvik@hig.no&view=cal