

Upscaling digitisation at the Wellcome Library showcasing the Goobi workflow system



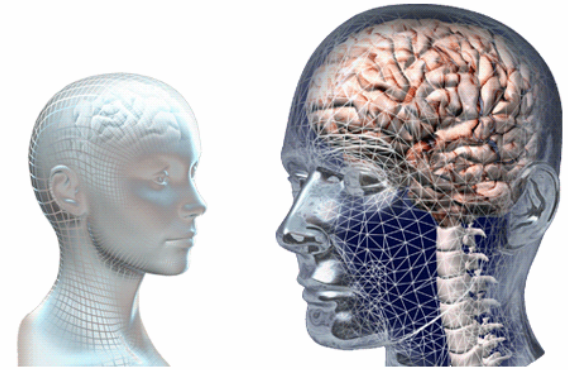
Christy Henshaw
Programme Manager
Wellcome Digital Library

3rd LIBER-EBLIDA Workshop on
Digitisation of Library Material in Europe

Koninklijke Bibliotheek
6 October 2011

wellcometrust

The Wellcome Trust



- A global charitable foundation
- Achieving extraordinary improvements in human and animal health
- Supporting the brightest minds in biomedical research and the medical humanities
- Exploring medicine in historical and cultural contexts



The Wellcome Library



- Major resource for the study of medical history
- Collections of books, manuscripts, archives, films and pictures on the history of medicine from the earliest times to the present day.
- Provide insight and information to anyone seeking to understand medicine and its role in society, past and present.
- Provide access to a growing collection of contemporary biomedical information resources relating to consumer health, popular science, biomedical ethics and the public understanding of science.

The Wellcome Library



Digitisation – the story so far



- Image library created from transparencies/prints, and on demand photography – 300,000 images
- Journal backfiles digitisation – (funder) Med. Hist., BMJ, etc. in PMC
- Wellcome Film – 500+ titles (also Wellcome Film YouTube channel)
- AIDS posters project – 3,000 posters
- Arabic manuscripts – 500 manuscripts
- 17th century recipe books – 74 manuscripts
- Contributions to Europeana via the Europeana Libraries project, and World Digital Library

The Library Transformation Strategy 2009 - 2014

To provide global access to, and expert interpretation of, a world class collection that explores medicine in its cultural contexts

- ***Targeted collecting*** – putting challenges in context
- ***Expert interpretation*** – engaging (new) audiences
- ***Strategic digitisation*** – online access to our collections



wellcome^{trust}

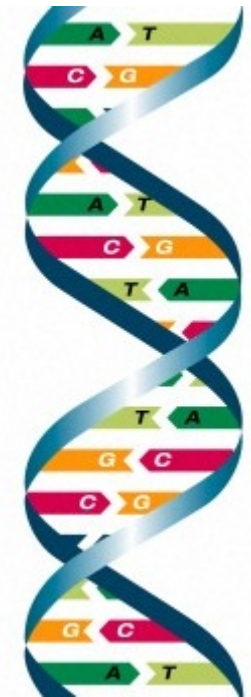
The Wellcome Digital Library pilot 2010-2013

Genetics and its Modern Foundations

A new online resource for everyone interested in the history of human and animal health.

Aims

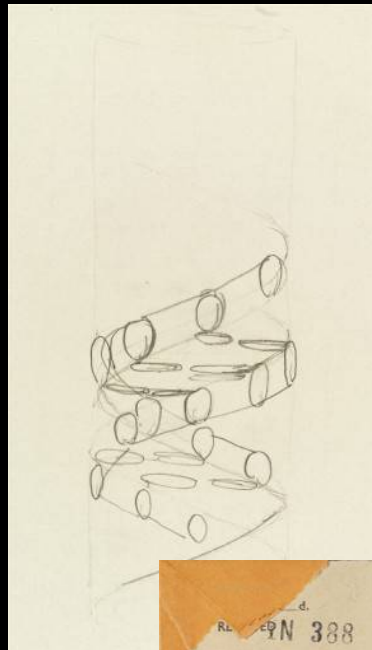
- ***build sustainable/expandable mechanism*** – foundation stone for WDL
- ***digitise key library holdings*** - relating to a major Trust challenge area
- ***digitise important third party content*** – linked to theme
- ***use innovative content and tools*** – to encourage discovery and use
- ***explore commercial partnerships*** – enhance access to non-theme material



Archival material – 900,000 images

Wellcome Library - 600,000 images

External – 300,000 images



POST OFFICE DL
TELEGRAM
Prefix, Time handed in, Office of Origin and Service Instructions, Words.
At 40 m At m
£ 216 GN L1484 STOCKHOLM 67W 18/10 1325
VIA NORTHERN =
DR FRANCIS HARRY COMPTON CRICK INSTITUTE OF
MOLECULAR BIOLOGY CAMBRIDGE
= THE CAROLINE INSTITUTE HAS DECIDED TO AWARD THIS
YEARS NOBEL PRIZE IN PHYSIOLOGY OR MEDICINE
JOINTLY WITH ONE THIRD EACH TO YOURSELF, DR
For full publication of complete words telephone TELEGRAMS ENQUIRY or call with this form to or C
at office of delivery. Other enquiries should be accompanied by this form, and, if possible, the envelope.

KEEP THE LEFTIES OUT: CRICK FOR GOD
ENOCH IS A RIGHT PRATT
1953 JESUS VOX
BUT WHAT IS STRANGE!

	U	C	A	G	
U	PHE PHE LEU LEU	SER SER SER SER	TYR TYR Ochr Anker	CYS CYS ? Tryp	U C A G
C	((leu)) LEU Leu	PRO PRO	HIS His	ARG ARG	U C
	leu (leu)	PRO PRO	GLU GLU	ARG ARG	A G
A	ILEU tLEU	THR THR	ASP ASP	(Ser) SER (Ser)	U C
	? ILEU MET	THR THR	LYS LYS	(Arg) ARG (Arg)	A G
G	VAL VAL	ALA ALA	ASP ASP	GLY GLY	U C
	VAL Val	(Ala) ALA ALA	GLU Glu	(Gly)	A G

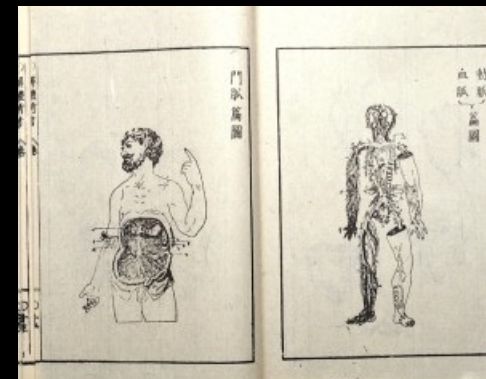
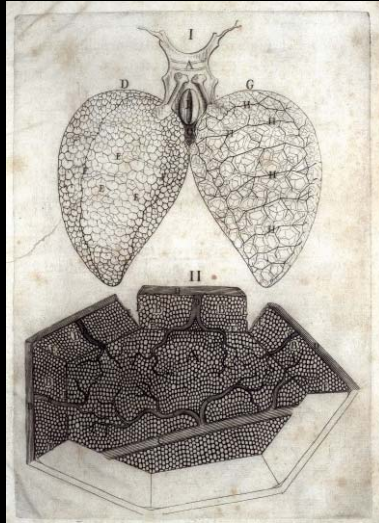
Capitol = Nuremberg, Nure
other = other Nuremberg results
and other Nuremberg

13th April '65
File

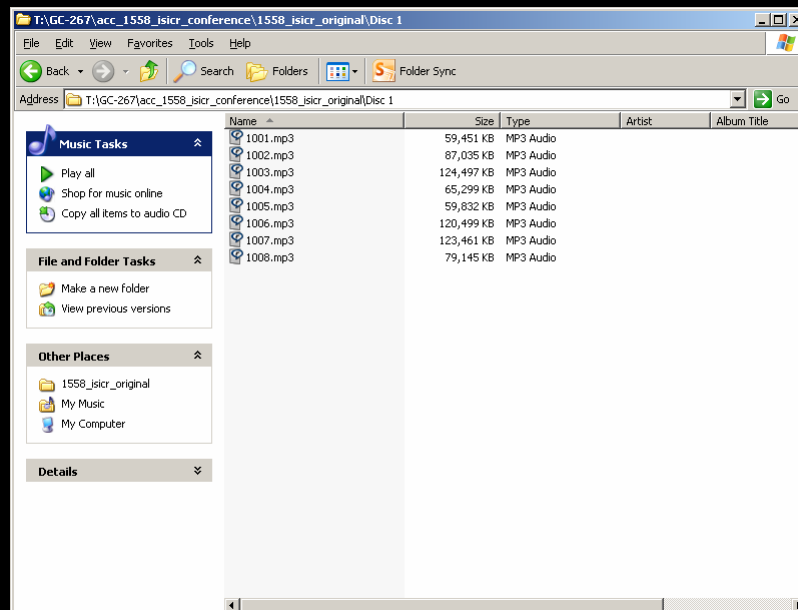
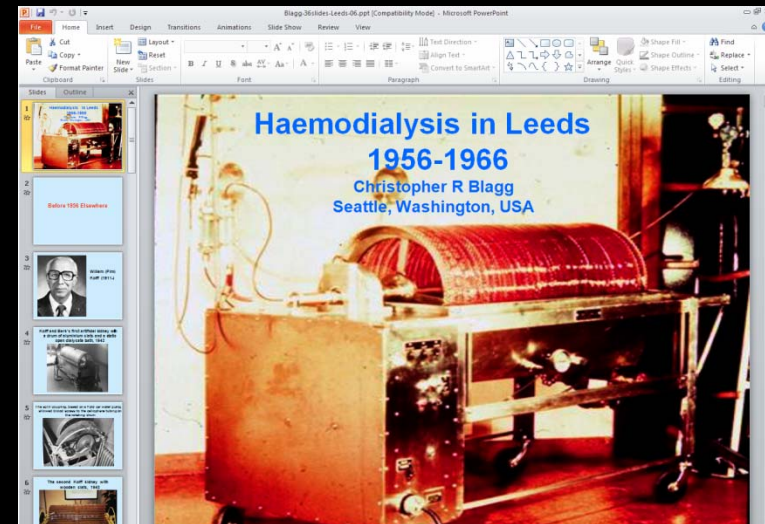
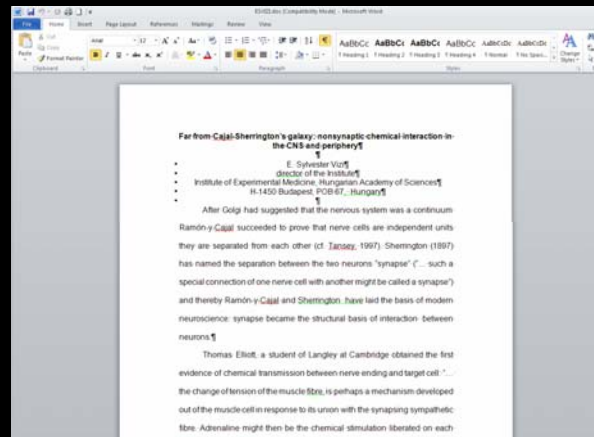
Books related to genetic research - 600,000 images



ProQuest, Early European Books – 5.5m images



Born digital material – initially small but growing




Digitisation strategy

Then	Now
Small projects (<10,000 pp)	Large projects (>100,000 pp)
Relatively ad-hoc	Major strategic programme
SMT & Project teams	Programme Board, advisors
Library-centric	W. Trust, external stakeholders
Entirely open access	Commercial partnerships
Little impact on IT systems	Requires major IT development
<i>Examples</i>	<i>Everything (within reason)</i>

Digitisation processes

Then	Now
Manual processes	Automated processes
Centralised conservation	Distributed conservation
Low QA	Increased QA, error minimization
TIFF	JPEG 2000
Individual tracking lists	Centralised tracking system
Incremental storage growth	Completely new storage strategy
<i>Detailed, painstaking</i>	<i>Streamlined, pragmatic</i>



Streamlining digitisation

- Staff dedicated to specific projects, or streams of work
- Carry out sample workflow tests for new types of material
- The right equipment for the right job – eliminate the “fiddly bits”
 - Live-view monitors
 - Easy-clean surfaces
 - Foot-pedals
 - Custom-made supports



Streamlining digitisation

- Photographers do the photography...
- Prepare materials separately
- Leave loose pages and bindings as they are, they are easier to digitise that way!
- Use existing staff as support – moving items to and from stack
- Minimise movement
- Keep plenty of shelving, working space at hand
- Find a preferred supplier for ad hoc support



Upscaling and streamlining digitisation
requires **a higher level of project
management**

Streamlining project management





What is it?

- Web-based workflow system
- Open source (core system)
- Used by many libraries in Germany, and half a dozen other European libraries
- **Intranda version** developed by Intranda to meet Wellcome Library specific requirements



What does it do?

- Task-focused, customisable workflows developed by Intranda
- User-specific “dashboard”
- Import/export and store metadata
- Encode data as METS
- Display progress of tasks, statistics on activities
- Tracks projects, batches, and units (location, current activity)
- “Command central” for 3rd party systems

<http://goobi.intranda.com>

intranda GmbH
digiverso

Log out

Home page

- General -
Technical background
Current users

- Workflow -
> My tasks

Search for volume
Batches

- User data -
User configuration
Change password

Log out:
Fürstlich, Wolfgang

Home page >> My tasks

My tasks

Please select one of the listed tasks.

Select pre-defined filter

Hits: 3

Show only my tasks:

Show open tasks only:

Filter:





Task	Process	Project	Status	Activity
<input type="checkbox"/> This is not a test, this is rock and/or roll <div> <div>Properties</div> <div> <div>Title:</div> <div>This is not a test, this is rock and/or roll</div> </div> <div> <div>Priority:</div> <div>0</div> </div> <div> <div>Last edited:</div> <div>Sep 7, 2011</div> </div> <div> <div>Last updated by:</div> <div>Engel, Detlev</div> </div> <div> <div>Update type:</div> <div>administrative</div> </div> </div>				

 a_ocm00665772i-dnt | Sample_Project | | || ☒ This is yet another test | a_ocm00665772ix | Sample_Project | | |
| ☒ Check images and import | _ThisisthePPNdigitalaSatz | Sample_Project | | |

Possible actions

Customise view

<http://www.goobi.org/> | intranda GmbH, Göttingen | Legal notice


<http://goobi.intranda.com>
intranda GmbH
digiverso



Log out

Home page

- General -
Technical background
Current users

- Workflow -
> My tasks
 Search for volume
 Batches










- User data -
User configuration
Change password

Log out:
Fürstlich, Wolfgang

Home page >> My tasks >> Details of the task


My tasks

Properties	
Title:	Check images and import
Process title:	_ThisisthePPNdigitalaSatz
Order:	2
Priority:	0
Status:	In process
Start of work:	09-Sep-2011 16:48:19
Last edited:	09-Sep-2011 16:48:19
Update type:	Manual, regular workflow










Paragraph
Font size

This book is very fragile, requires post-digitisation preservation.


Path: n


Send correction message to previous task




Back to task

Remark

Send correction message


Finish this task

Possible

-  Re
-  Se
-  Fir

<http://www.goobi.org/> | intranda GmbH, Göttingen | Legal notice

Project management tasks

intranda GmbH
digiverso

Log out

Home page

- General -

Technical background

Current users

- Workflow -

My tasks

Search for volume

> Current processes

Batches

Create new process

- User data -

User configuration

Change password

Log out:
von Darst, Cecilia

Home

Pr

Cui

Cre

Hits

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

a

Possible actions

- Set status of process up
- Set status of process down
- Calculate number of metadata and images
- Statistical evaluation
 - Volume status
 - Responsible user groups
 - Duration of the steps
 - Production statistics
 - Production throughput
 - Storage calculator
 - Error tracking
 - Project association
- Generate excel

Statistical evaluation: Error tracking

Define statistical question

Period of bis

unit View

Select pre-defined filter

show finished processes: ☐ Filter:

Project	Activity
Sample_Project	
Sample_Project	
Sample_Project	
Sample_Project	

en | Legal notice



Administrative tasks

Goobi intranda GmbH Log out

http://goobi.intranda.com

Hits: 5

show archived projects

Process title ▼▲

- BasicWorkflow
- Born_digital_material
- External_digitisation
- Inhouse_digitisation
- Large_Workflow

Number	Title	Status
1	Base information	
2	Barcode generation	
3	DMD import	
4	Image capturing	
5	Automatic QA	
6	Qualitative QA	
7	DAM DMD export	
8	DAM AMD data import	
9	OCR	
10	AccessCondition and Copyright	
11	Data export to DAM	

Activity

Log out: von Darst, Cecilia

Digital asset management

File conversion

- Lightroom used to convert RAW to TIFF
- LuraWave converts TIFF to JP2K
- Validation of JP2K conversion coming soon – *via Goobi*

Ingest

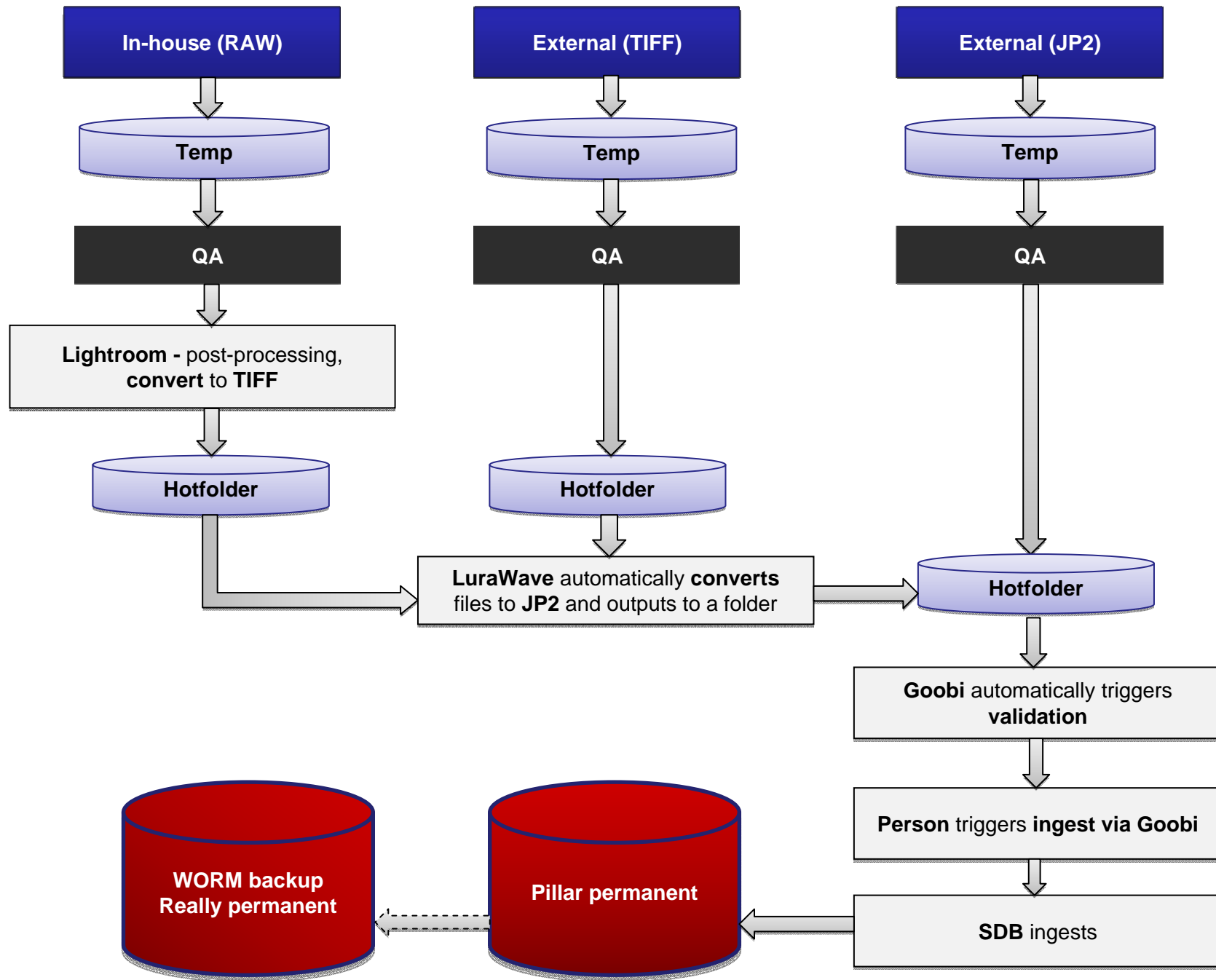
- Automated ingest workflow in the DAM (Safety Deposit Box - SDB) – *via Goobi*
- One file serves as master and dissemination file

Preservation

- DAM is a preservation system
- Manages all preservation actions (characterisation, format migration)
- API to allow 3rd party systems access to content

Storage

- Master files backed up offsite to WORM storage drive
- WORM = Write Once Read Many – permanent storage
- Self-healing of errors on main storage system from WORM



Thank you!

Christy Henshaw

c.henshaw@wellcome.ac.uk