

# Upscaling digitisation at the Wellcome Library showcasing the Goobi workflow system

**Christy Henshaw**  
Programme Manager  
Wellcome Digital Library

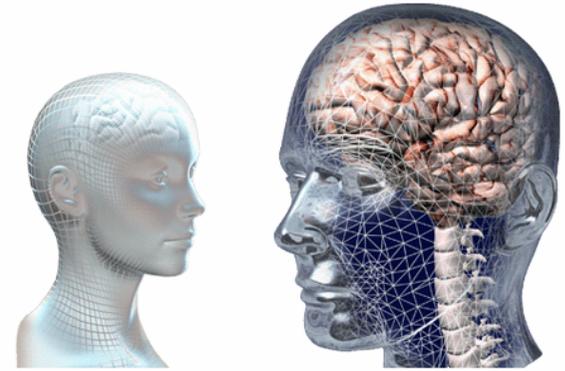
3<sup>rd</sup> LIBER-EBLIDA Workshop on  
Digitisation of Library Material in Europe

Koninklijke Bibliotheek  
6 October 2011



**wellcome**trust

# 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
- 17<sup>th</sup> 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



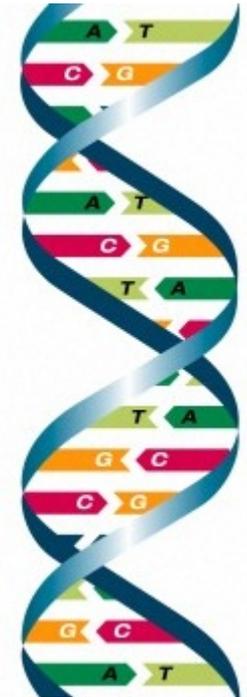
# The Wellcome Digital Library pilot 2010-2013

## *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



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

POST OFFICE DL OFFICE STAMP 18 OCT 62

TELEGRAM

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

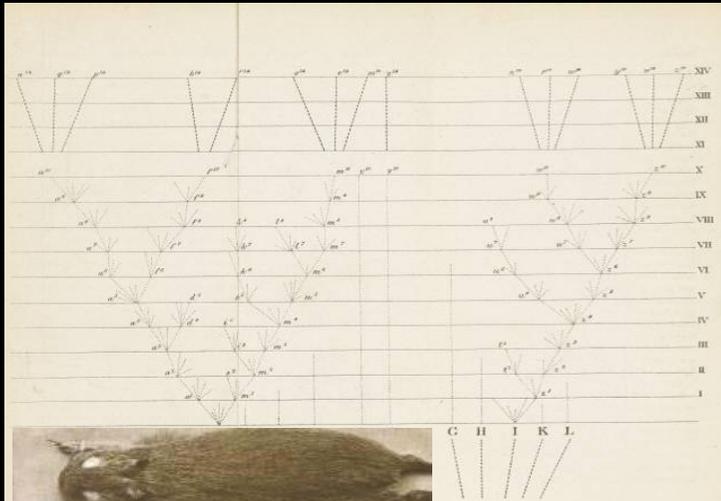
JOINTLY WITH ONE THIRD EACH TO YOURSELF, DR

	U	C	A	G	
	PHE	SER	TYR	CYS	U
U	PHE	SER	TYR	CYS	C
	LEU	SER	Ochr	?	A
	LEU	SER	Ambr	Tryp	G
	((leu)) LEU	PRO	HIS	ARG	U
C	Leu	PRO	HIS	ARG	C
	leu	PRO	GLN	ARG	A
	((leu))	PRO	GLN	ARG	G
	I LEU	THR	ASP	(ser) SER	U
A	I LEU	THR	ASP	(ser)	C
	? LEU	THR	LYS	(arg) ARG	A
	MET	THR	LYS	(arg)	G
	VAL	ALA	ASP	GLY	U
	VAL	ALA	ASP	GLY	C
G	VAL	(Ala) ALA	GLU	(gly)	A
	Val	ALA	Ser		G

Capitals = Nucleic acids  
 others = other various results and other sources.

15 April '65  
 File

# Books related to genetic research - 600,000 images



C H I K L

so extraordinary or so beautiful a manner, and are furnished with such curious and elegant crests of hair, that we can hardly avoid looking at these characters as having been gained for the sake of ornament. The accompanying figures (figs. 72 to 76)



Fig. 73. Head of *Semnopithecus comatus*



Fig. 74. Head of *Cebus capucinus*.



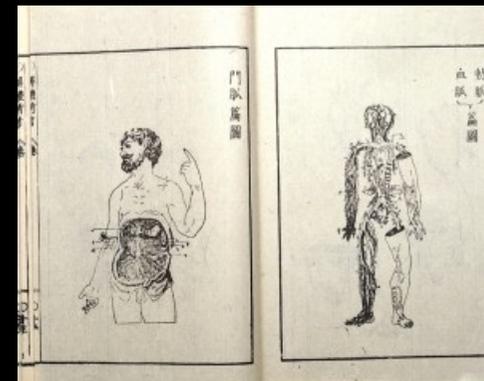
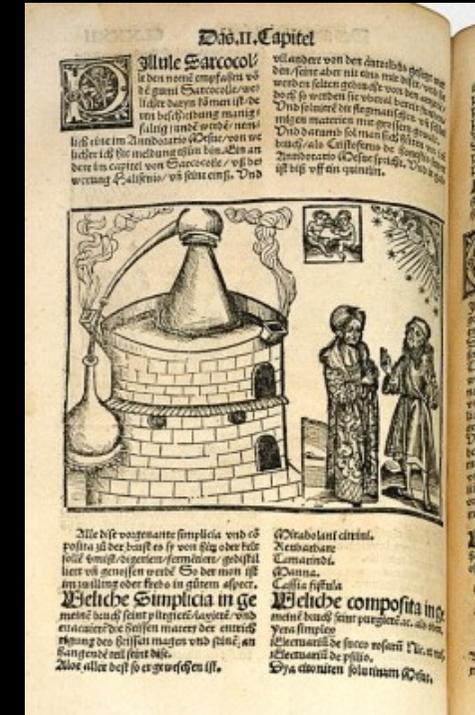
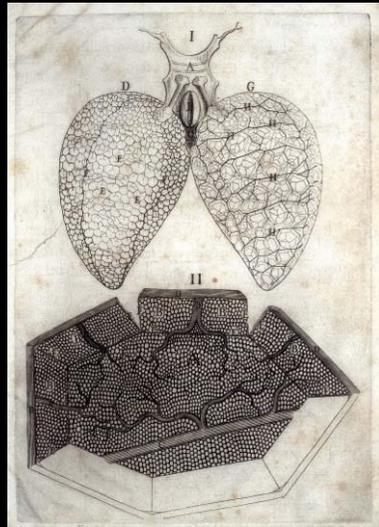
Fig. 75. Head of *Ateles marginatus*.



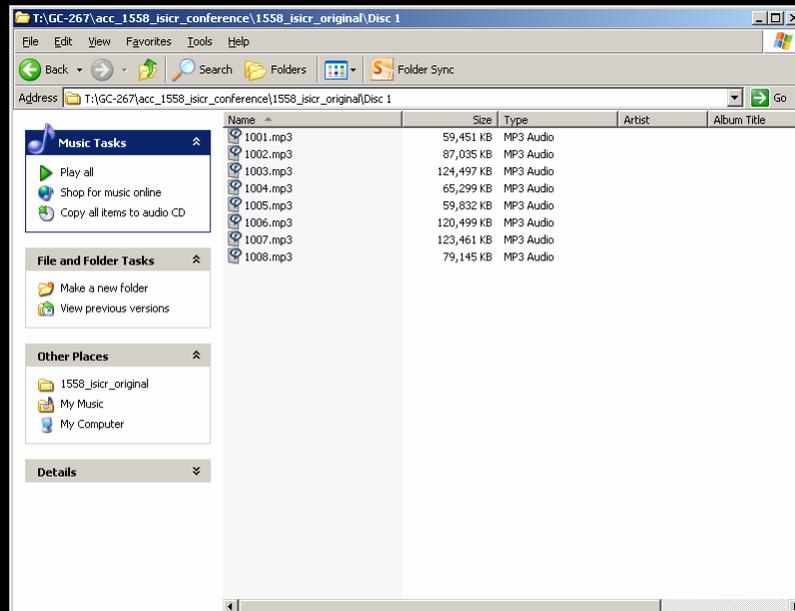
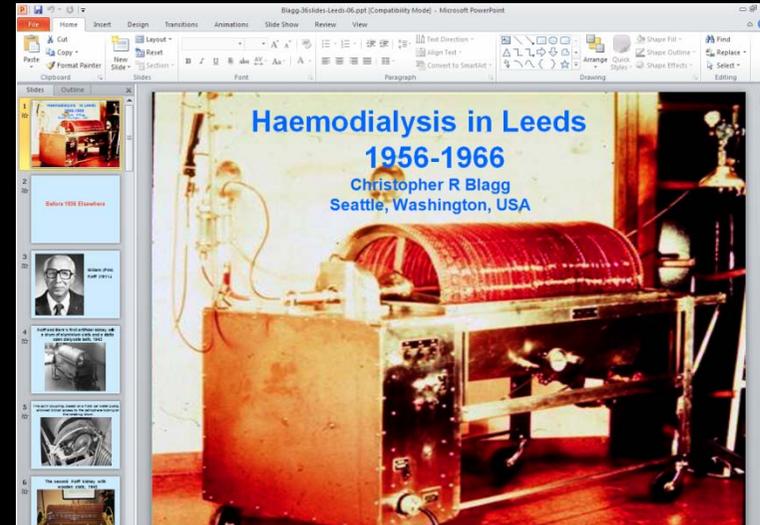
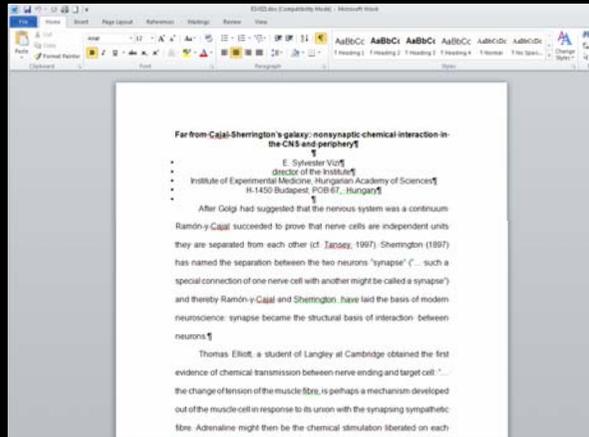
Fig. 76. Head of *Cebus vellerosus*.



# ProQuest, Early European Books – 5.5m images

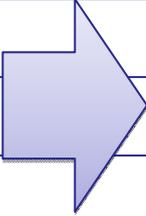


# Born digital material – initially small but growing



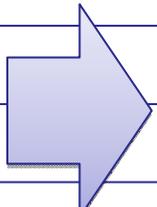
# Digitisation strategy

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
<b><i>Examples</i></b>	<b><i>Everything (within reason)</i></b>



# 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
<b><i>Detailed, painstaking</i></b>	<b><i>Streamlined, pragmatic</i></b>



# 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 IntraData
- 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 3<sup>rd</sup> 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
<div style="border: 1px solid #ADD8E6; padding: 5px; margin-bottom: 5px;"> <p><input type="checkbox"/> This is not a test, this is rock and/or roll</p> <p><b>Properties</b></p> <p>Title: This is not a test, this is rock and/or roll</p> <p>Priority: 0</p> <p>Last edited: Sep 7, 2011</p> <p>Last updated by: Engel, Detlev</p> <p>Update type: administrative</p> </div>	a_ocm00665772i-dnt	Sample_Project		
<input type="checkbox"/> This is yet another test	a_ocm00665772ix	Sample_Project		
<input type="checkbox"/> Check images and import	_ThisisthePPNdigitalaSatz	Sample_Project		

Possible actions

Customise view

<http://www.goobi.org/> | intranda GmbH, Göttingen | [Legal notice](#)

intranda GmbH  
digiverso

  
Log out
 

http://goobi.intranda.com

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

**B** *I* U ABC | 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

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



intranda GmbH  
digiverso
 


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

+ \_

Pos

⋮

⋮

📊

📊

**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	📄 📄 📄 ⬇ ⬆ 📄 METS PDF DMS
➤	Sample_Project	📄 📄 📄 ⬆ ⬆ 📄 METS PDF DMS
➤	Sample_Project	📄 📄 📄 ⬆ ⬆ 📄 METS PDF DMS
➤	Sample_Project	📄 📄 📄 ⬆ ⬆ 📄 METS PDF DMS

# Administrative tasks

Goobi intranda GmbH Log out

http://goobi.intranda.com show archived projects

Hits: 5

Process title ▼▲

- BasicWorkflow
- Born\_digital\_material
- External\_digitisation
- Inhouse\_digitisation
- Large\_Workflow

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

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

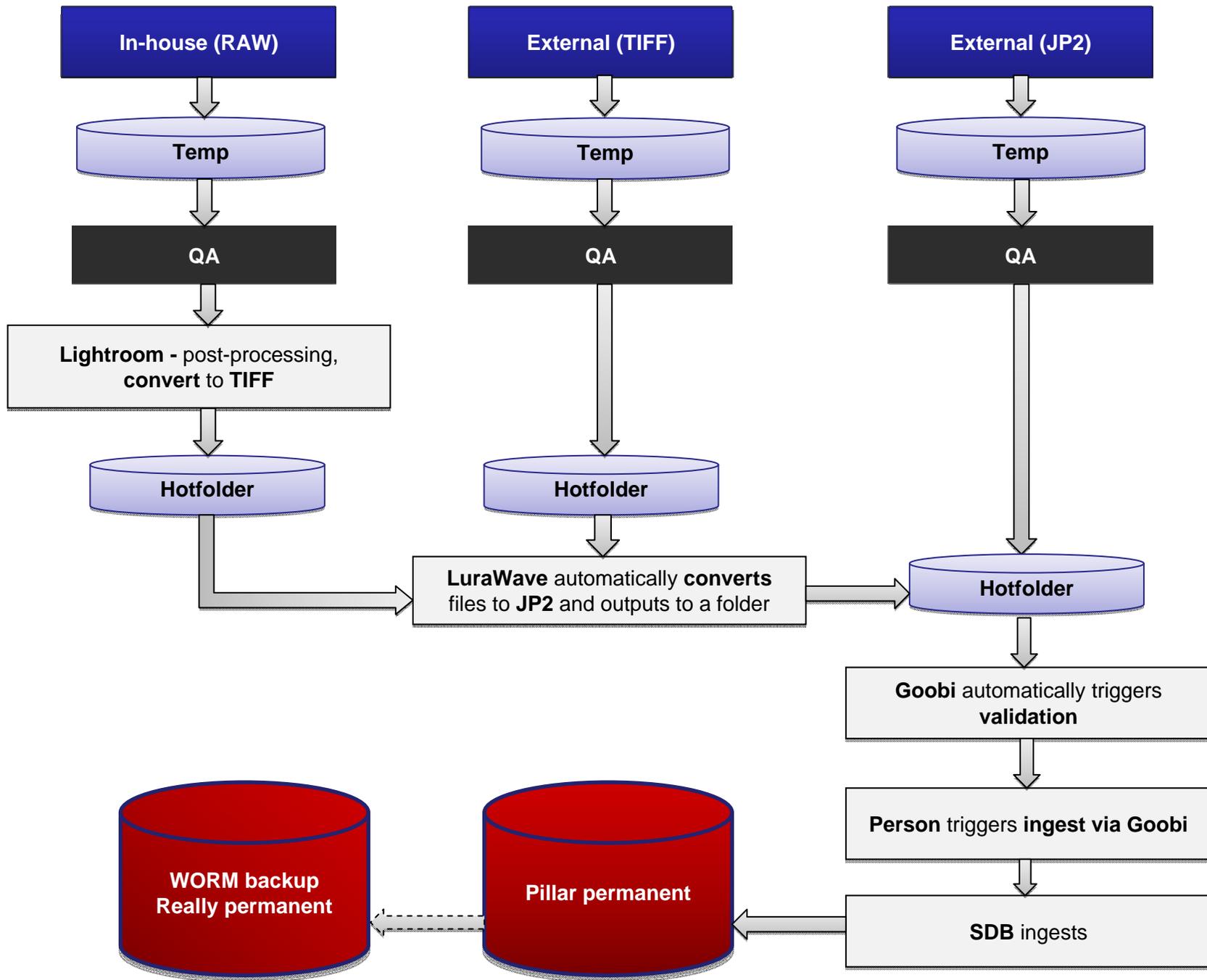
- Automated ingest workflow in the DAM (Safety Deposit Box - SDB) – *via Goobi*
- One file serves as master and dissemination file

## Preservation

- DAM is a preservation system
- Manages all preservation actions (characterisation, format migration)
- API to allow 3<sup>rd</sup> 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](mailto:c.henshaw@wellcome.ac.uk)