



CENDARI: What Next?

Dr Jennifer Edmond
Unique and Universal: Challenges for the
Manuscript Librarian
25th May 2016









Presentation Structure

- What is CENDARI?
- Three perspectives on 'next'
 - For the current CENDARI (sustainability)
 - For the next CENDARI (systemic alignment)
 - For future CENDARIs (new paradigms)
- Concluding thoughts







FP7 Integrating Activity for Infrastructure

14 Consortium Partners, 8 Countries 2012-2016, total budget €6.5M

- Freie Universität Berlin
- University of Birmingham
- Trinity College Dublin
- Czech National Library
- Università di Cassino
- The European Library
- Consortium of European Research Libraries
- King's College, London
- INRIA
- University of Stuttgart
- Goettingen State and University Library
- Mathematical Institute of the Serbian Academy of Sciences and Arts
- Fondazione Ezio Franceschini
- Società Internazionale per lo Studio del Medioevo Latino









Primary Aim of CENDARI

Pilot a research infrastructure leveraging analogue networks to integrate digital resources for historical research in a way that is easy to use/access and harmonised with researcher requirements

Two Pilot Areas

- World War I
- Medieval European Culture











Resources, Resources

- Lists of institutions, catalogues, finding aids
- Search and metasearch engines
- Databases (authority lists, standardised author names, shelf marks)
- Dictionaries
- Manuscript collections: ISOS, Manuscriptorium, Monasterium, eCodices,, Gallica, Mandrigoria...but all in silos

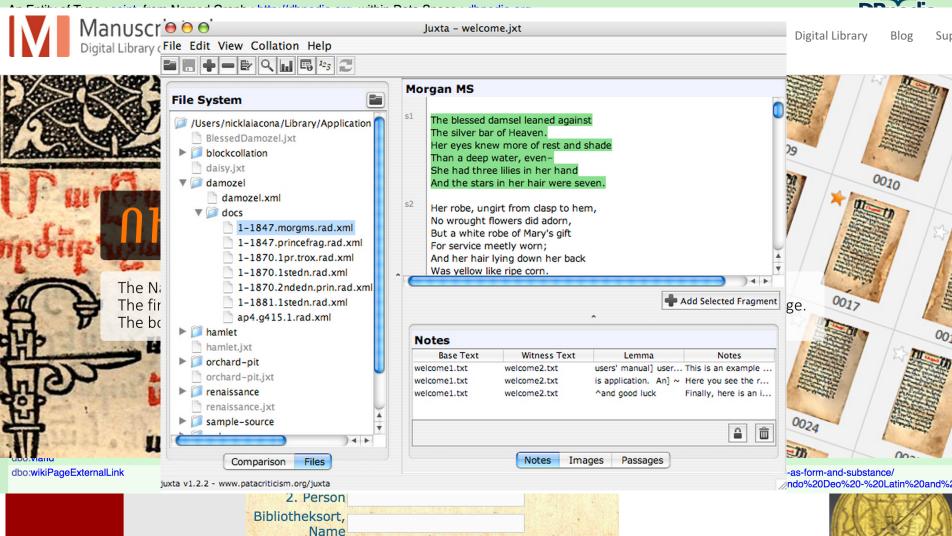
- Tools, to let you do cool things (virtial vellum, hyperimage, transcriptoria)
- Juxta, Collex, reading environments for annotation and parallel versions
- Projects and networks: more important as a group of people (COST IS 1005), Leeds/ Kalamazoo
- MESA, platform for resources to be peer reviewed and validated





About: Albertus Magnus





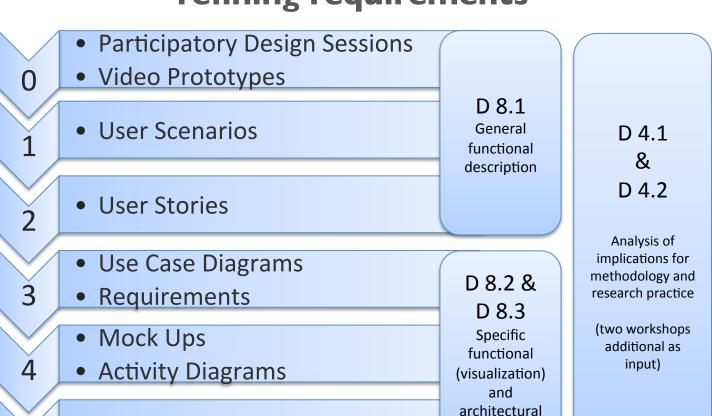
logische Verknüpfung der Begriffe

...but how? Infrastructure!



Prototypes

Process of understanding researchers and refining requirements



description





0a. Participatory Design Sessions



Paris, October 2012:WWI researchers

Florence, January 2013: Medievalists

Berlin, October 2013: Archivists, Librarians and Information specialists

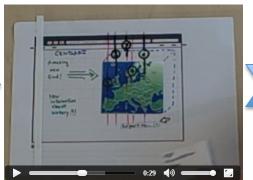






Ob. Result of Participatory Design Sessions: Video Prototypes









• Mapping of documents and files













I. User Scenarios

"My project examines how the rural-urban divide shaped Habsburg Austrian society's experience of the war from about 1915 (when food and food shortages became increasingly politicized) [...]

I want to answer the following questions: How did the administration and realities of rationing vary between cities on the one hand, and between urban centers and the rural areas of their provinces on the other?

In terms of research, I will first need to map the changing structures of rationing and the incidence of food protests and related disturbances over the course of the war. [...]

In the event that I locate a particularly rich vein of primary sources [...] I would be interested in digitizing them and subjecting them to text-mining analysis. Specifically, I could measure the changing occurrence of terms like "country" in the discourse of city-dwellers and "city" or "town" (or the name of a city) in rural discourse. [...] relative occurrence of such terms against other dividing categories such as the "nation" or "the rich." [...]"

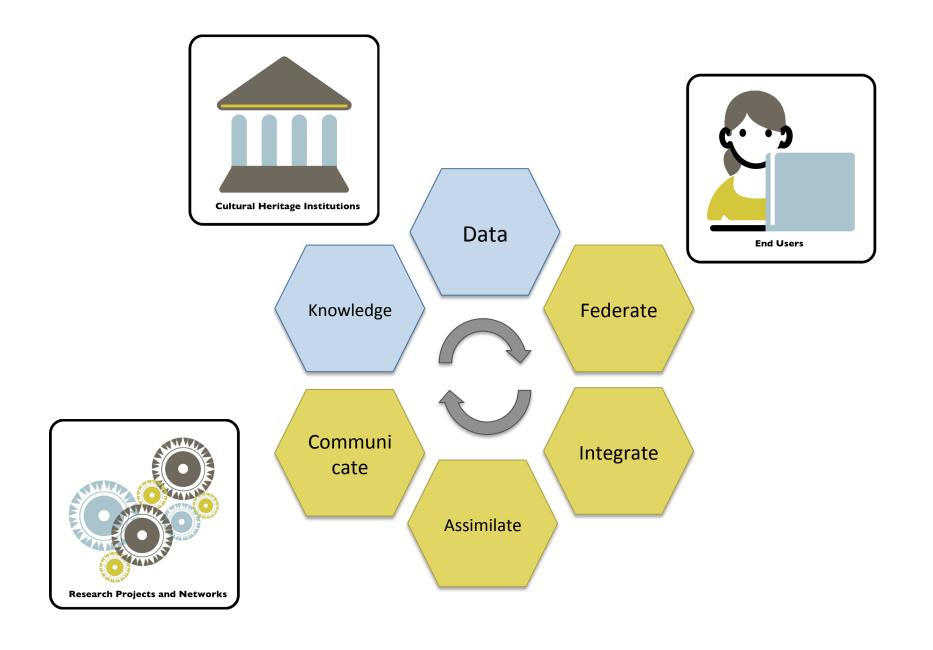




2. User Stories

User Story Id	As a (role)	I want to (something)	So that (benefit)	Reference to Source
US-1.1	researcher	View and browse the stacks in "closed stack" collections	I can discover valuable materials that I would not have thought to search for specifically.	"Shelf Screenshot" Video Prototype
US-1.2	researcher	Search for archives and information about their history in a particular city or region	I can do research on a certain city or region	"Archival Network by Location" Video Prototype
US-1.3	researcher	Visualize my searches	I can see which archives I have visited and how often, and where they are located	"Visualizing Search Paths" Video Prototype
US-1.4	researcher	Search for translations/transcriptions of documents in CENDARI	I can see what translations/transcriptions have already been done by other CENDARI users	"Crowdsourcing Translation of Documents," part 1 Video Prototype
US-1.5	researcher	Work with other researchers online to translate/transcribe a document	I can create a better translation/transcription	"Crowdsourcing Translation of Documents," part 2 Video Prototype
US-1.6	researcher	View/manipulate/navigate documents (enlarge, reduce, rotate, etc.) and select/highlight parts of documents and add annotations	I can create translations/transcriptions linked to the original document	"Crowdsourcing Translation of Documents," part 2 Video Prototype
US-1.7	researcher	Search for archives by type in a particular location	I can see what kinds of archives are available	"Geo-Search" Video Prototype
US-1.8	researcher	View documents and files according to chronology (temporal), using a timeline or location (spatial), using a map	I can understand and display the spatial or chronological relationships between documents	"Mapping of Documents and Files" Video Prototype
US-1.9	researcher	View a 3D projection of documents that are mapped over both time and location (both chronological and spatial)	· · · · · · · · · · · · · · · · · · ·	"Mapping of Documents and Files," Part 2 Video Prototype



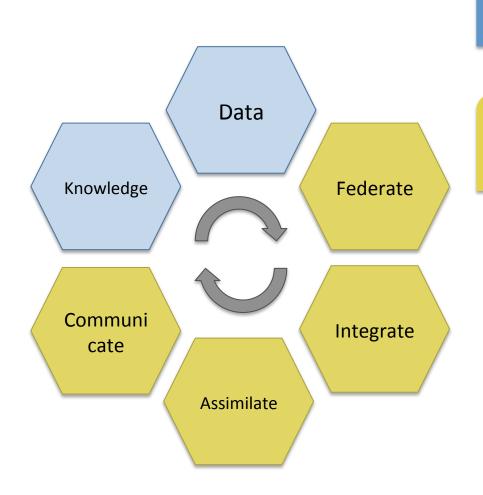


Challenges faced by archives in sharing data

'Crisis' of digital sustainability and reuse

Tensions in the publication environment

Subtlety of current practices (travel)



Participatory
Design
Methodology

'Data Soup' and hidden archives, frustration with siles

LOD and ULS approaches

Note-taking as a key mode of processing

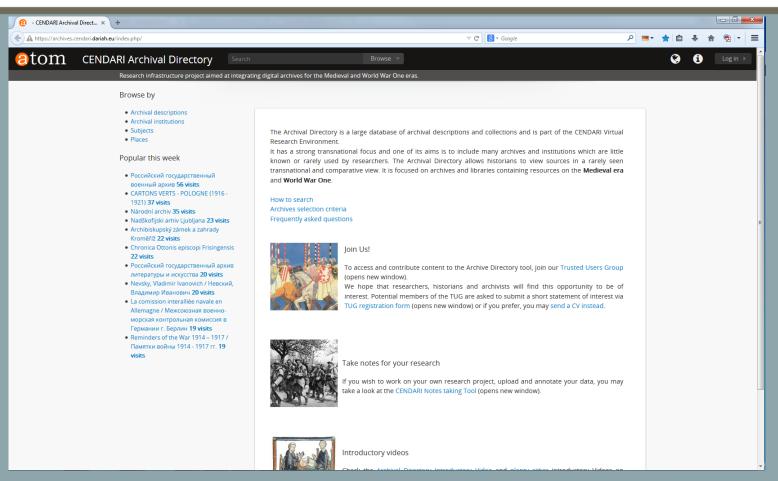


What does the system do (for a user)?

- Support 'enquiry' phase
 - Search, Browse, Add Collections
 - Personal workspaces: combine system information with your own archival records, transcribe, annotate...
 - Explore themes through the Archival Research Guides







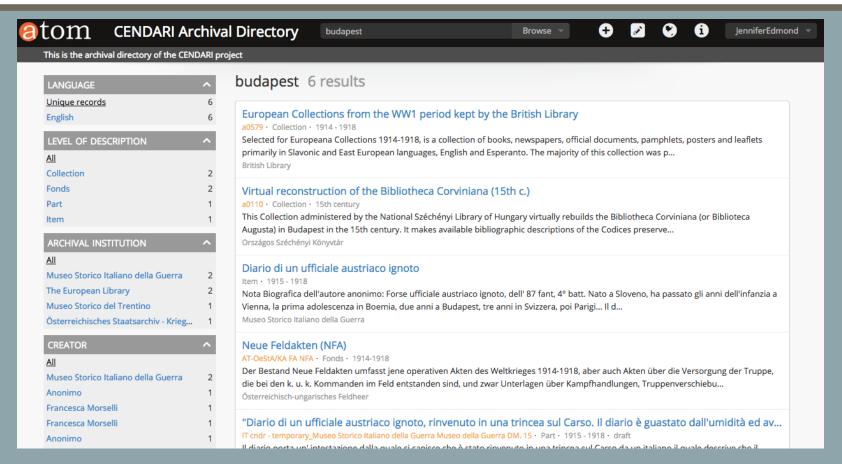
Archival Directory AtoM





Cultural Heritage Institutions in the Archival Directory





CENDARI Repository CKAN

Currently 303,225 datasets, and more in process

Search the records:

https://archives.cendari.dariah.eu/index.php/?module=repository



Archive of the British School at Athens

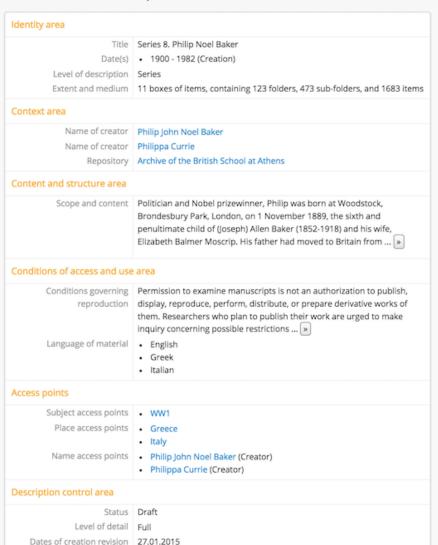
Holdings Quick search

▼ Series Series 8. Philip Noel Baker

Subseries SUB-SERIES 1.1 - FAMILY

Searching the records:
Collection Level

Series - Series 8. Philip Noel Baker



Reports

Export

Dublin Core 1.1 XML

EAD 2002 XML

Related subjects

WW1

Related people and organizations

Philip John Noel Baker

(Creator) Philippa Currie (Creator)

Related places

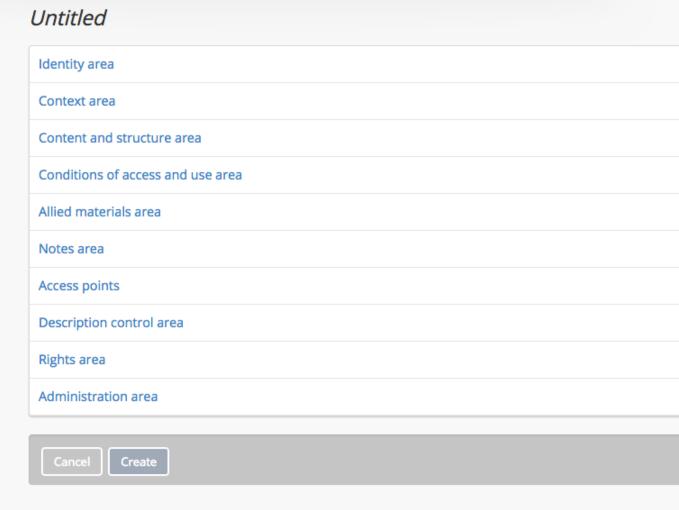
Greece Italy





This is the archival directory of the CENDARI project

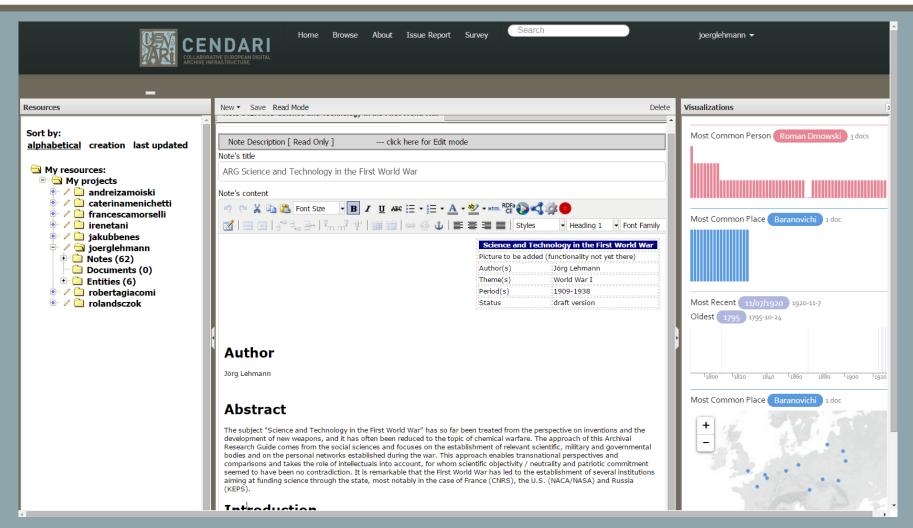
Manual Creation of a Collection Description



page for the creation of the Institution (EAG):
https://archives.cendari.dariah.eu/index.php/repository/add
page for the creation of the collection (EAD):
https://archives.cendari.dariah.eu/index.php/informationobject/add







CENDARIS Note Taking Environment (NTE) and ARGS

https://notes.cendari.dariah.eu



Ţ	Complete Menu - 1	Home / About & Partner	s •⊃Login •	■ Register	⊠ Contact Us				
	Search with Trame								
	_								
	▼ Set your search								
	Browse by list Se	earch Packages			Selected sources				
	Is FIRB? Y	Select Souce Type:	rchivio 🗆 Bibliote	ca	Registerd users can save search configuration and load search packages.				
	Show 10 centries		Search:		Log in Register				
	Check All _ Type	Name			Perform your Search				
	_ *	BAI (Mirabile) Biblioteca Agi	ografica Italiana						
		BDC: Beinecke Digital Collec	ctions						
		Biblioteca Apostolica Vatica	na Manoscritti Digi	talizzati					
		Biblioteca Apostolica Vatica	na – Catalogo dei n	nanoscritti					
		Biblioteca Medicea Laurenz							
		Biblioteca Pinacoteca Accad	lemia Ambrosiana						
		Bibliothèque nationale de F	rance – Gallica, ma	nuscrits					
	· •	BISLAM (Mirabile) Bibliothe Recentiorisque Aevi	ca Scriptorum Latir	norum Medii					

TRAME II

http://trame.fefonlus.it/trame/index.html





Methodologies emphasizing

ARCHIVAL

Research Guides

Examples:

- Parallel Records to Reconstruct Lost Material
- Medieval Libraries:
 Dispersion and
 Survival

Archival Research **GUIDES**

Archival

RESEARCH

Guides

Examples:

- RomanovDynasty
- Vernacular
 Bibles

Examples:

- Coercion and Consent; POW
- Prophecy and Political Thought in the Medieval Age





CENDARI
COLLABORATIVE EUROPEAN DIGITAL
ARCHIVE INFRASTRUCTURE

Home

About

AtoM

Issue Report Surve

Searc

jenniferedmond -

Table of contents (show)

The aim of this Archival Research Guide is assist the user when reading primary sources and transcribing them into their own notes onto the Virtual Research Environment platform of CENDARI. It complements the many thematic medieval Archival Research Guides already uploaded by researchers onto the system. As this guide is focused on online resources, the lack of stability of such resources must be stressed: Links might be broken, software might be non-compatible, etc.

Introduction

Abstract

Palaeography is the study of ancient handwriting. Letter forms make up a 'script', a particular type of handwriting for a certain place and time. The way in which each individual scribe writes a script is called a 'hand'. Hands can be divided into 'book' and 'documentary' hands, depending on the source. Codicology is the study of the codex, and examines the book as a physical object and how it was produced. Objects of study could include writing material (parchment, paper), ink, binding, the folding and division of the leaves, page layout, the illumination, the book's 'history' (ownership), marginalia, etc. Details of all of these concepts and their presentation in primary source materials can be found in the digital resources presented in this guide and in the bibliography. Nomenclature often varies from one author to another. Palaeography is an essential skills for medieval scholars, as nearly all of source material predates the invention of printing.

Note of the author

The digital resources listed in this Archival Research Guide were selected as relevant sources for training and teaching transcription. It is aimed to assist researchers who are starting to work on original documents. A few relevant resources for early modern historians have also been included. This guide does not aim for completeness and does not cover all digital resources on manuscript transcription currently available. Future contributors are encouraged to add new digital resources if they cover the following subjects: palaeography, codicology, digitisation of medieval manuscripts, and medieval book history. Similarly, if the digital resource linked to the ARG is no longer extant, contributors may delete the entry.

Palaeography Tutorials

There are various online tutorials and training materials available to improve palaeographical skills. They are also helpful if the researcher wishes to practice transcribing digitised documents.

Tutorials for medieval material

Tutorials on Medieval Palaeography (developed by the University of Leicester, now hosted by the Anglo-Norman Online Hub)
Medieval Writing

Course in Latin and Syriac palaeography and transcription, Hill Museum and Manuscript Library (also see site for a lexicon and annoted manuscripts for training)

The Medieval Handwriting App (developed by the University of Leeds)

Collaborators Theme(s) Medieval Culture Period(s) Status Final version Collaborators Silvia Nocentini Theme(s) Medieval Culture Period(s) Status Final version

Transcribing Medieval Manuscripts and Archival Material

ARG Transcribing Medieval Manuscripts and Archival Material

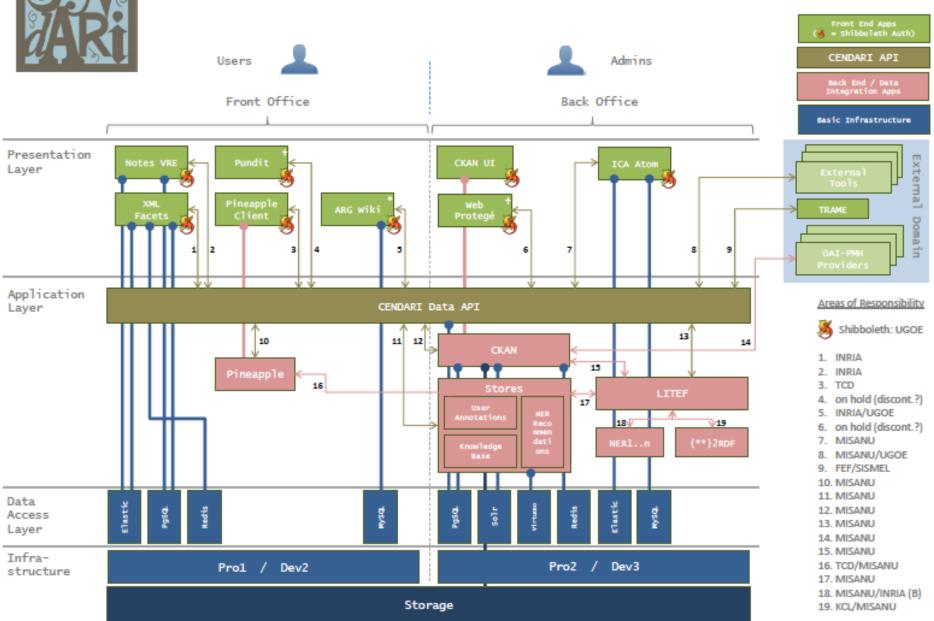


What else does the system do?

- Ingest and harmonise data and descriptions from heterogenous sources without investment in new metadata description
- Utilise a central data API to coordinate between data sources and services
- Self-monitor for changes in content and apply basic transformations
- Harness LOD to create intelligence across the system



CENDARI Infrastructure Layer Model v0.7



CENDARI October 2014



What else has the project done?

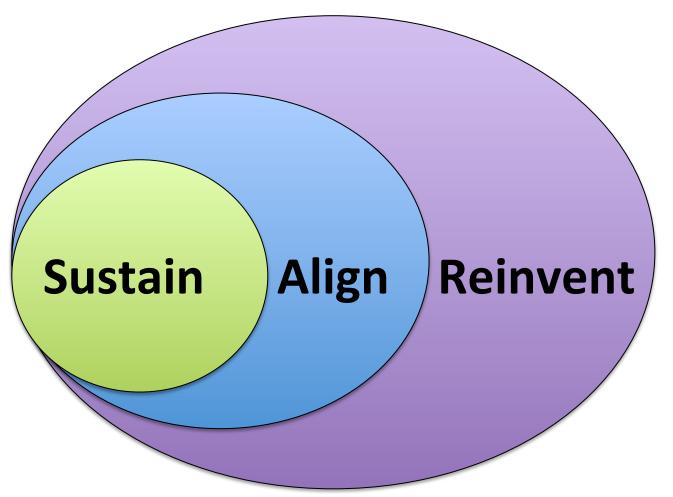
- CENDARI promotes systemic change in the following areas
 - Development of instruments for coordinating between these levels
 - Promote change in publication patterns
 - Working with DARIAH to embed new perspectives on sustainability of digital projects







So, what next?







The CENDARI Sustainability Model

- Many models for sustainability, each dependent on a project's self image
 - An organisation (sustain people, funding)
 - A technical infrastructure (sustain software)
 - Data (sustain bits, access)
 - A user community ('commons' or 'foundation' model)





Reuse = Sustainability

- Research infrastructures are not digital libraries
- Most RI initiatives are not optimised for the long term
- DARIAH provides an exception to this rule, and an opportunity for a different approach to sustainability





CENDARI's 'Asset Register'

- Technical Infrastructure: Portal, Services (VM) and Tools
- Research Data: Unique and Aggregated
- Internal Publications and Knowledge: Scientific Publications (ARGs) as well as 'Tacit Knowledge'
- Communities: People, Networks and Relationships





CENDARI's Recommendations for Future Projects

- 1. Proposals should provision a sustainability process, including (at minimum) a midpoint meeting of stakeholders and agreed final plan
- 2. DARIAH affiliated projects should have DARIAH representation/participation at VCC level
- 3. Projects should use common platforms and ...
- 4. apply known and accepted standards in their work
- 5. Project publications should cover scientific, technical and management aspects of the project, with different audiences in mind for different pieces
- 6. Every project should have a data management plan
- 7. Projects should be able to guarantee the availability of their basic technical formats for 3 or 5 years after close
- 8. Successor funding projects should build, rather than just maintain, open up to new user groups and/or new functionalities





What would CENDARI do next?

- 6. Lobby for greater acceptance of non-traditional publications
- 5. Open up the API
- 4. Improve look and feel of the NTE
- 3. Complete work on LOD elements (reduce up-front investment in description)
- 2.
- 1.





What would CENDARI do next?

2. Work to understand historians better

- CENDARI's robust participatory design and development process was not fundamental enough
- How do we meet a desire for 'completeness'
- How do we support non-developers to understand what is at stake in their methodologies?





What would CENDARI do next?

1. Work to understand CHIs better

- How to harmonise national/local contexts with transnational requirements?
- How to assist with resource barriers?
- How to harness immense tacit knowledge?
- How to build trust? How to harmonise policies so as to facilitate reuse?
- CENDARI 'White Book' and the Open History report on 'Sustainable Digital Publishing of Archival Catalogues'





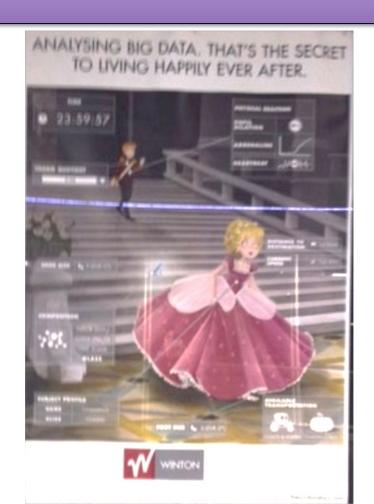
What does the future hold?

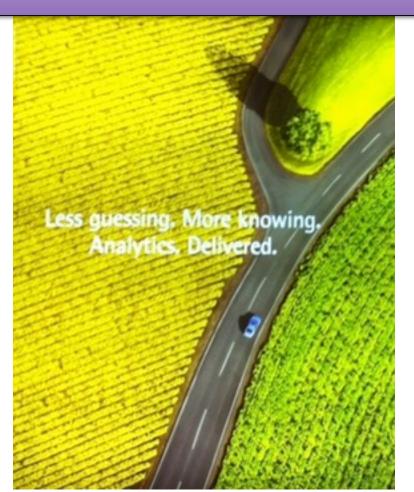
- 1. On the one hand, digital data needs to be open to individual or community remix and curation (eg. through open APIs)
- 2. One the other hand, we MUST become better at expressing the uniqueness of humanities research and cultural data





Big Data is Everywhere...







How do we get to 'big enough' data?

- Better collaboration and communication between computer science, information science and historians, involving interaction and dialogue, but also self-reflection.
 - How can we better express what the knowledge creation process for the humanities is?
 - Why are there summer schools for humanists to learn python, but none for computer scientists to learn humanistic approaches?
 - How do we deal with the imperative for open science when we have entire disciplines with no raw data?



How do we get to 'big enough' data?

- New mental models: for example what would a hybrid system for historical research (rather than only a system where all content is digitised) look like?
 - Google Streetview? IOT? Aircraft autopilot?
- What might 'big enough' data look like?
 - Above the project level
 - Beyond linked open data (dictionaries and concordances)



How do we get to 'big enough' data?

- Build systems that support trust.
 - Content holders need to trust that new pathways for the use of their materials will not lead to the exploitation of individuals or of the resources themselves
 - Researchers need to feel confident in collections and their ability to use them
 - Scholars need trust systems to make their underlying limitations and assumptions transparent

