



Digital Research Infrastructure
for the Arts and Humanities

Digital infrastructure at the service of open science

Laurent Romary

Inria, Directeur de Recherche

DARIAH, director

www.dariah.eu

THATCamp@Göttingen

Overview

- Going open, an epistemological stance
 - Publications, data, open culture, barriers
- The open science agenda of DARIAH
 - Showing the way at European level

Context

- Quite some progress on open access to publications
 - Better awareness of the difficulties encountered with the current publishing ecology
 - Increasing role of publication repositories — cf. TARA at TCD, ORBI in Liège, HAL in France etc.
 - Towards new publication models (from blogs to epi-journals)
- Sharing data in the humanities is not the mainstream (yet)
 - We need to understand why in order to take action
- Scholarly reluctance, lack of recognition, missing technical infrastructure?

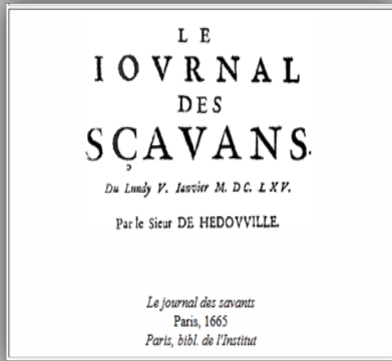


Digital Research Infrastructure
for the Arts and Humanities

Scientific publications where are we?

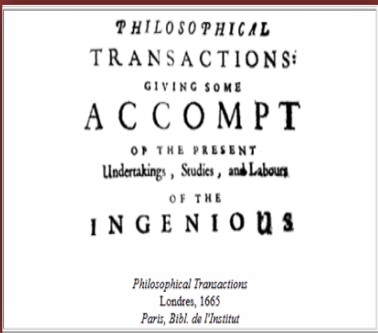
www.dariah.eu

The scholarly publishing system: the origins « 1665 »



Journal des sçavans

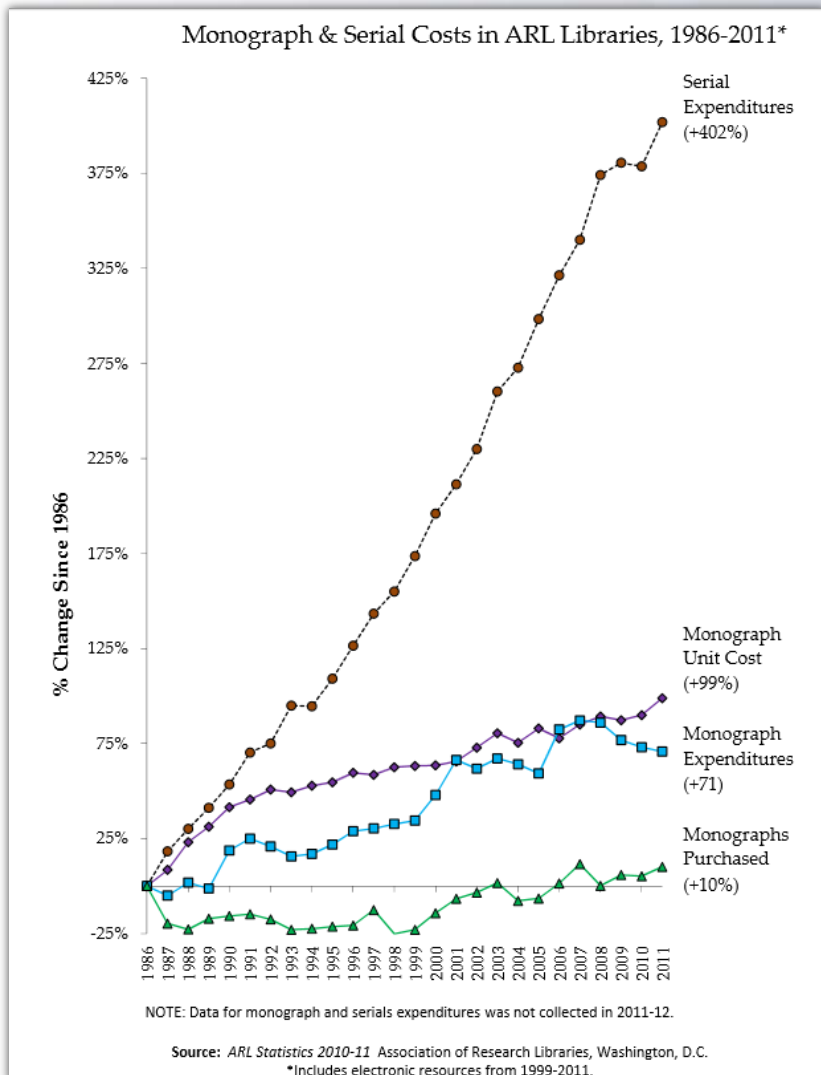
- Spread & communicate scientific discoveries
- Dissemination & archiving on a **stable medium**



Philosophical Transactions of the Royal Society of London

- Established the principles of **peer-review** and scientific priority
- Genesis of **scientific communities** (scientific networking)

The serial crisis



- ✓ Dramatic increase of subscription cost in American libraries
400% between 1986 and 2011
- ✓ Journals became the subject of a **very lucrative international business** of which libraries are **captive**

Profits by Elsevier

- 2002: £429m profit on £1295m revenue – 33.18%
- 2003: £467m profit on £1381m revenue – 33.82%
- 2004: £460m profit on £1363m revenue – 33.75%
- 2005: £449m profit on £1436m revenue – 31.25%
- 2006: £465m profit on £1521m revenue – 30.57%
- 2007: £477m profit on £1507m revenue – 31.65%
- 2008: £568m profit on £1700m revenue – 33.41%
- 2009: £693m profit on £1985m revenue – 34.91%
- 2010: £724m profit on £2026m revenue – 35.74%
- 2011: £768M profit on £2058M revenue – 37.3%
- Want to buy shares?
- Source: <http://svpow.com/2012/01/13/the-obscene-profits-of-commercial-scholarly-publishers/>

Customers are producers

Learned societies
& Publishers

Researchers and
institutions are
creating value

Research

Are they providing
good services for a
fair price?

Article writing

How much do they
work at the service
of research?

Reviewing
Evaluation process

Open access doxa

- Berlin Declaration (2003)
 - „...free, irrevocable, worldwide, right of access to, and a license to copy, use, distribute, transmit and display the work publicly and to make and distribute derivative works, in any digital medium for any responsible purpose....“
- Green open access — Author’s deposit within a publication repository
 - Which version? When? Under which control?
- Gold open access — Online access provided by publishers (possibly against an article-processing-charge APC)
 - Native open access, hybrid models, freemium model (OpenEdition)
 - Publishers tend to preserve their revenue
 - Very costly for RPOs (Research Performing Organisations)

An exemplary policy: Inria

- Full green
 - Deposit mandate for all publications in HAL
 - Condition for all reporting activities
 - Exploring new models: the Episciences project
 - Using publication repository as the basis for new journals
 - Combining publication and data-sets
- Towards a cautious gold
 - Hybrid payments are forbidden
 - Centralised budget and monitoring of all APCs

Hurdles remain to be overcome

“All rights reserved. Without limiting the rights under copyright reserved above, no part of this book may be reproduced, stored in or introduced into a retrieval system, or transmitted, in any form or by any means (electronic, mechanical, photocopying, recording or otherwise) without the written permission of both the copyright owner and the author of the book.”

Real example taken from a PhD thesis from XX



Digital Research Infrastructure
for the Arts and Humanities

Are we ready to go open access
with our data?

www.dariah.eu

The quite usual case

- Gisela Musterfrau has made a beautiful transcription of an 19th century manuscript
- She puts it online as an HTML page on her google+ account
- She's now wondering if she did it right...
- Well no, she was afraid that the content could be stolen
 - So she still puts the final touch on it...

Gisela could have done better

- Authorship
 - Who is responsible for the work online
 - E.g. providing feedback
- Format
 - The document can only be read, not re-used
 - No semantics attached, even if all proper names have been put in italics...
- Technical sustainability
 - No trustful digital setting
 - What if the google account is being blocked
- Re-use instructions
 - Two ways of doing things wrong
 - No clear license attached to the document
 - Is the reader allowed to take up even a paragraph?
 - Reserving all rights

Where Laudatio did it right

- Authorship
 - Clear identification of data sources in the Laudatio data model
- Format (standards)
 - A comprehensive (meta-)data interchange format based upon the TEI guidelines
 - Publication of all specifications and associated documentation
- Technical sustainability
 - Trustful digital setting at CMS
- Re-use instructions
 - Clear re-use instruction by means of the CC-BY licence

Is it wise to openly share Humanities data?

- Why would it be unwise?
 - Data can be “stolen” by other scholars
 - Is there a risk that other scholar carry out the same research?
 - People could sell the data, publish it?
 - WTF, as long as it does not prevent data to be freely available elsewhere
- A sensible issue: digital editions, in depth annotations
 - Early scholarly recognition
 - Competition with traditional publication means
 - Scholars still promise a book in their project applications
- Is it wise not to openly share humanities data?
... whenever it is possible

What benefits accrue from open access to Humanities data for researchers, and for citizen?

- Pooling scholarly results
 - Avoid duplication of work on primary sources
- Comparing and checking result
 - Towards more evidence in the humanities
- Showing that hard work is being done
 - Cf. more visibility to humanities research in the wide public
- Humanities data are cultural heritage data
 - General public interest

The possible role of a research infrastructure

What is DARIAH?









- An integrating activity bringing together digitally-enabled research in the arts and humanities in Europe
- A platform enabling trans-national arts and humanities research
- A research infrastructure for sharing and sustaining digital arts and humanities knowledge
- By researchers, for researchers








A quick history of DARIAH

- **2006:** DARIAH included in the ESFRI Roadmap
- **2008 – 2011:** Preparatory Phase project, *Preparing DARIAH*
- **2011 – 2013:** ‘Transition Phase’, establishing the DARIAH-ERIC, integrating national activities
- **2014+:** DARIAH-ERIC



15 Committed Founding Members

Founding Members	
Austria	
Belgium	
Croatia	
Cyprus	
Denmark	
France	
Germany	
Greece	

Founding Members	
Ireland	
Italy	
Luxembourg	
Malta	
Netherlands	
Serbia	
Slovenia	

Candidate countries	
Lithuania	
Poland	
Portugal	
Spain	
Switzerland	

A complementary view on open scholarship: DARIAH short-term goals

- Inventory
- Standards
- Certification
- Legal issues
- Service sustainability

How does this help?

- Inventory
 - More visibility, dissemination of simple entry points (cf. URL in Annis!)
- Standards
 - The Laudatio data model as a contribution to community standards
- Certification
 - Providing trust to the Laudatio repository (e.g. Data Seal of Approval)
- Legal issues
 - Recommending CC-BY more widely
- Service sustainability
 - A correlate to openness
 - Providing long term access to existing corpora

Overview

- Contributing to an open science culture
 - At individual level
 - The central role of senior scholars: panels, personal choices
 - At project level
 - Showing good practices: licenses, standards, accessibility, sustainability
 - At community level
 - Open science as part of the agenda of digital infrastructures



DARIAH-EU

Digital Research Infrastructure
for the Arts and Humanities

Merci !

www.dariah.eu