Elements of a scientific communication policy

Laurent Romary

Elements of ~~an Open Access~~ a scientific communication policy

Laurent Romary
Inria, scientific information advisor

PROLOGUE: A MINED LANDSCAPE

A vivid actuality

Good news?

- Julia Reda and the reform of copyright law in Europe
  - But there is no implementation yet…
- Open science as a major theme of the Dutch presidency (January-June 2016)
  - "All European scientific articles to be freely accessible by 2020"

Bad news?

- SSRN and Hivebench, following Mendeley a couple of years ago, now in the hands of Elsevier
- OA2020 initiative by the Max Planck society…
  - "Transition" to which model?

OA battle won by STM publishers?

Nobody will benefit if a major European industry
is undermined and with it the peer review system
upon which science and society depend,
February 2006

STM press releases over the years…

Open access (OA) is most commonly defined as making
original research freely accessible on the web, ideally
immediately on publication. STM publishers are currently
either implementing, or testing, ways to harmonise this
visionary goal with economic reality.
April 2008

'Gold' open access publication is the practical route to achieving sustainable
open access, the project partners agreed today at the PEER End of Project
results conference in Brussels.
May 2012

Inria

The French resistance? ;-)

Geneviève Fioraso, Ministre de l'Enseignement supérieur et de la Recherche (24 Jan. 2013):

"Scientific information is a public good that should be available to all"

- A policy considering all aspects of scholarly publishing
    - Author's manuscript deposit, Author-pays models, Alternative models
- A policy based on the support to scientific information infrastructures
    - HAL (Publication repository), OpenEdition (Digital publishing platform), BSN (Bibliothèque Scientifique Numérique)

Scientific legal setting: Loi pour une République Numérique (2016)

- Principle of author's deposit (well... with embargoes)
- TDM rights for legaly acquired scholarly material

Inria

Identifying what we should be acting upon

Taking distance with events big and small

Three main action lines

- Controlling the budget dedicated to third party services
  - Should 1% of the world research budget be dedicated to printing flyers?
- Defining the practices that are relevant for scientific communication
  - Are we happy with the current journal system (peer-review, citation index, lack of dynamicity, Baroque technological management)?
- Ensuring our digital sovereignty
  - Who should take care of our scientific commons ?

Overview

The development of a scientific information policy at Inria

How to become colour-blind

Identifying what we have in stock

A pragmatic construction of a scientific information infrastructure

INRIA: TAKING OUR DESTINY IN OUR OWN HANDS

A favourable background and a vision

4,400
People
(60 % paid by Inria)

8 Research Centers in France

41 International conferences

66

Scientific publications 4,450

180 Project teams

250 Active patents

1,000 Doctoral students
100 Post-Doctoral
300 R&D engineers
3,500 Scientists

A BUDGET OF
€ 265M
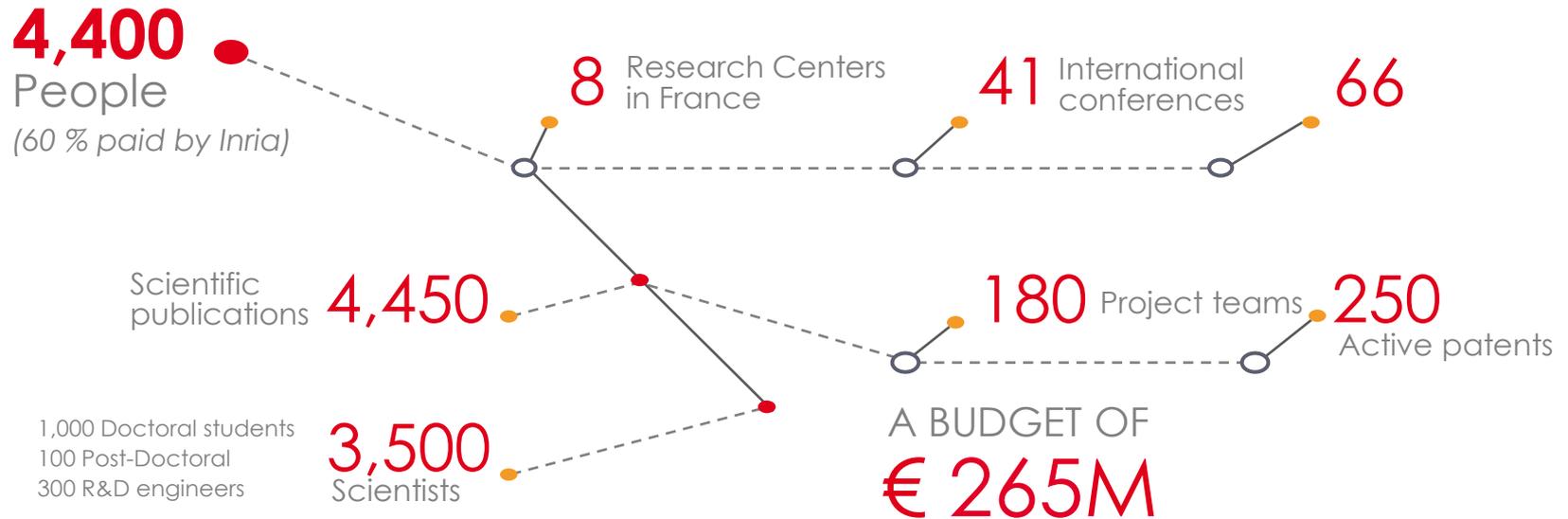
Our priorities for a scientific information policy:

- Maximal dissemination of our scientific assets (visibility and swift dissemination of knowledge), for a reasonable price

- Constitution of a reliable and sovereign institutional corpus (documentation, preservation, access), with clear public governance principles

- Contribution to shaping the scientific communication landscape in terms of editorial processes and with regards the usage made of scientific productions

Inria

Green open access as a baseline

- Deposit mandate
  - A prerequisite for the annual assessment of research teams

- Benefits for the researcher
  - Providing an increased visibility to his/her research productions

- Benefits for the institution
  - Creating an exhaustive corpus reflecting Inria's presence
  - Informational sovereignty

- Accompanying services
  - National network of librarians checking and enriching the deposits

Issues — 1

A deposit mandate should not come as a surprise

- Strong recommendation policy since 2005

Coverage – a culture of openness

- From early manuscripts to post-review documents

Licences – going for attribution (CC-BY)

- Favouring the fluid dissemination of knowledge

Affiliations and authorities

- Persons, teams, institutions, projects
- Essential for developing precise exploration tools

Issues — 2

Involvement of library staff

- Digital curators at the service of the Inria information space

Assessment

- Seen as a service by our researchers
- Goes with the development of tools for creating personal or team websites, multiple export formats

Embargoes

- Not mentioned in Inria's policy…

Deposits on the Inria portal
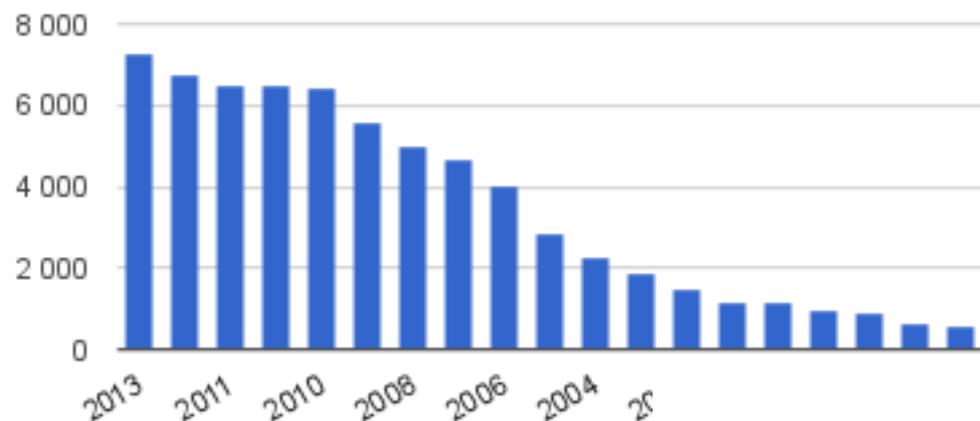
Per year

Cumulated:
• Blue: total number of entries
• Yellow: full-texts
• Green: bibliographical records

61400 entries as of Oct 2014
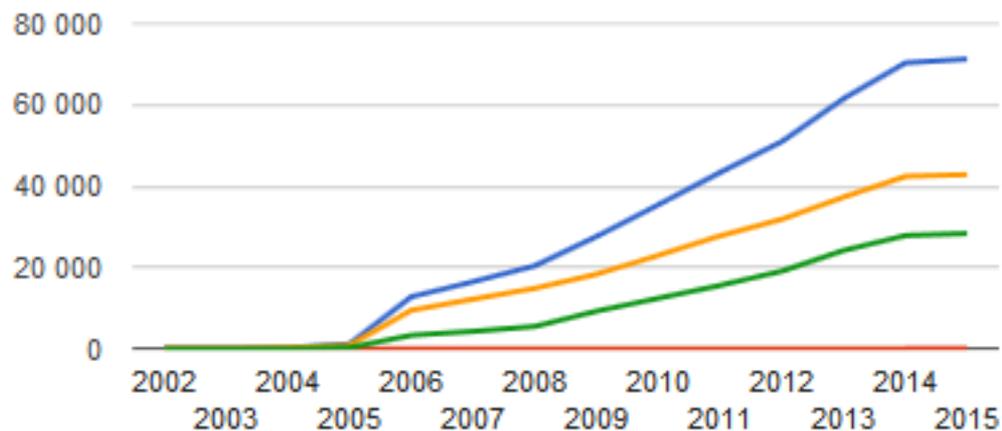
Full text content in the Inria portal
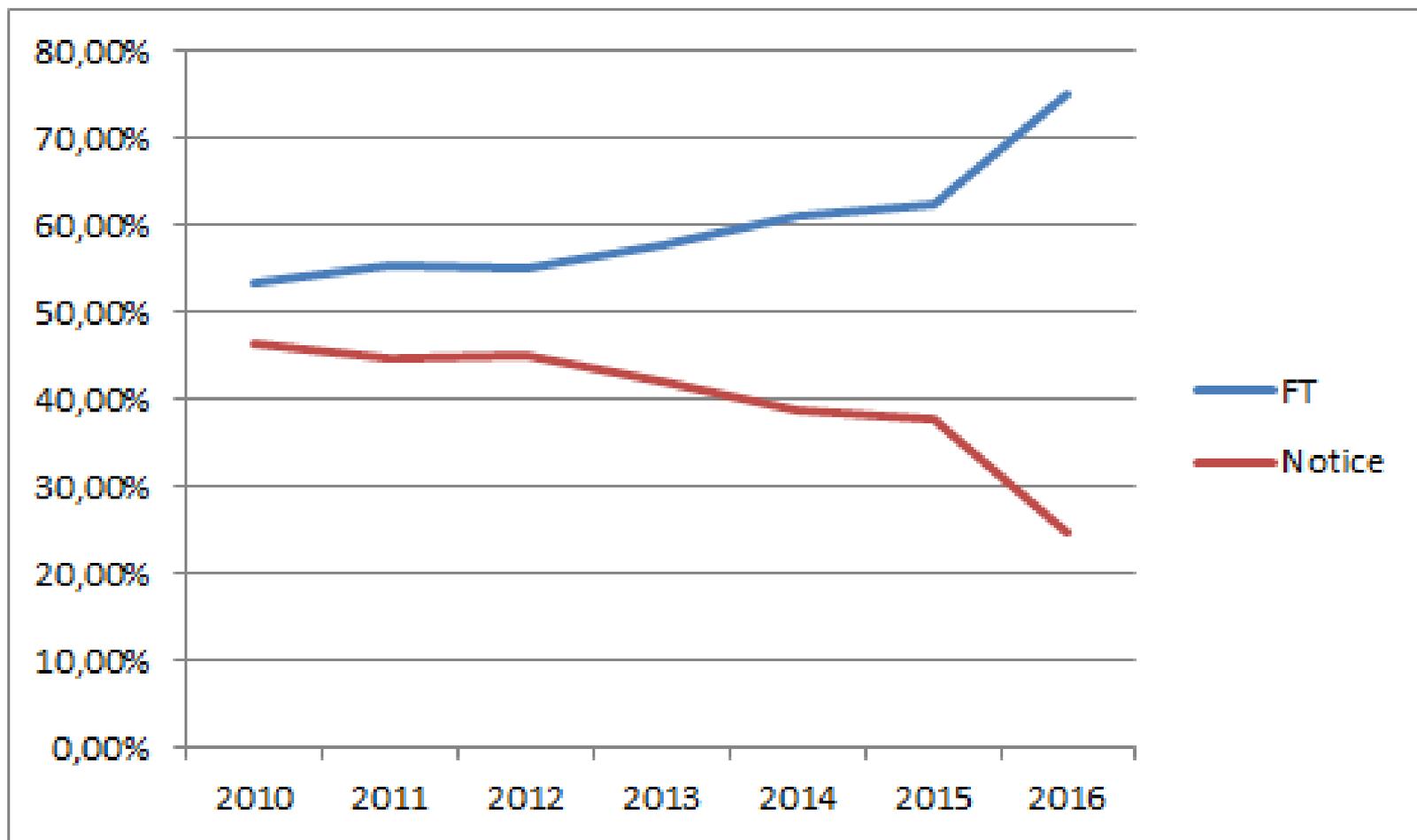
The essential role of authors' manuscripts

Early scholarly dialogue

- Communicating results to receive feedback
- Versioning as part of the communication process

Marking up anteriority

- Time stamping of results to prevent plagiarism and state anteriority
- http://www.google.com/patents/EP2547095A1

A step towards making publication more autonomous from certification

- A continuum from colleagues' feedback to community review
- Peer review as the certification of a specific version

Keeping Gold under control

- Why being cautious?
  - Because we are a research performing organisation
  - Because we have a deposit mandate

- Banning the hybrid model

- A centralized support to Gold open access
  - Single and open-ended (!) budget
  - Precise documentation of use
  - Information about costs and predatory publishers

- Gold means green
  - Registering gold papers in HAL (with identical licence)

Ínría

ARE WE FOOLED BY COLOURS?

The green road

- Essentially locked in its relation to traditional publishing mechanisms
- "peer reviewed journal articles": there is an outer force that determines what should be communicated and we need to set up a "parallel" dissemination mechanism
- No need for any other model? The system will toggle automatically, like a piece of magic, when…
- Consequence: outer categories
    - pre-prints, embargoes (sign of an external temporality: the reference is the time of "publication" in official journals)
- Scope reduction of the content: author's manuscripts (not "published"), stop calling them pre-prints…
    - slides, posters, outreach and pedagogic materials… theses

The golden road

- Even worse
- Alternative financial model for exactly the same journals: "transition"
- No guaranty of budget reduction
- Opening up a predatory market
- Which room is left for other models (Episciences, Freemium)?
- Debate on open licences fouled by the publishers' take up of it
- The private sector has played the open access card well… (cf. CC licences)

Green and gold as conservative categories

Scientific information policy

- Series of measures associated with an institutional vision
    - Excellence in the creation of scientific knowledge
    - Transfer to the civil and economic society
- Not the desire to implement colours, nor to do "open access"

The green/gold opposition does not favour the implementation of deep changes in the scientific communication landscape

TAKING STOCK

Starting with what we have: HAL

Initiated by the CNRS in 2001 as a mirror to ArXiv

- Quick expansion to all scientific domains, comprising human and social sciences
- Increasing support from higher education and research institutions
- 360 000 full-text documents in open access
- Among which 34 000 doctoral theses
- More than 3 000 new documents every month
- A joint endeavor of CNRS, Inria and Universities

A few available benefits

Authors' services

- Personal identifier (IdHAL) and web pages (CVHAL)
- Collections and portals

Authorities

- AureHAL: authors, institutions, journals, projects

Licences

- Associating a deposit with a variety of possible licences

Interfaces and formats

- Standards ingestion (SWORD) and export protocols (OAI/PMH)
- Relying on a highly structured open format: TEI (Text Encoding Initiative)

Mutualisation

- Central development, distributed editorial management
- Ex. Grobid

Going beyond the limitations of pdf

publishing →

← text-mining
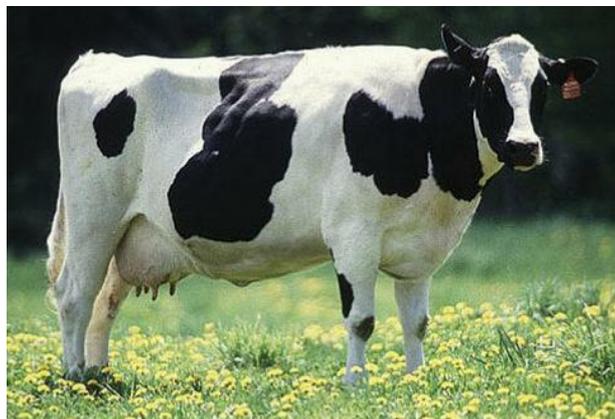
Cow (structured data)

Hamburger
(unstructured data)

"Converting PDF to XML is a bit like converting hamburgers into cows. You may be best off printing it and then scanning the result through a decent OCR package."
Michael Kay (http://lists.xml.org/archives/xml-dev/200607/msg00509.html)

Inspired from: Duncan Hull

Structuring content

- GROBID: information extraction from PDF documents
  - Meta-data: title, authors, affiliations, abstracts
  - Bibliographical references (with crossref consolidation)
- Standards base representation
  - TEI (Text Encoding Initiative) as a reference format
- State of the art quality (CRF models)
  - Cf. M. Lipinski, et al., 2013
  - Used at EPO/ResearchGate/Mendeley
- Integrated in HAL
  - Automatic meta-data extraction for author's deposit
- On-going developments
  - Full-text extraction
  - Anticipating the re-publication of content (HTML, ePub, etc.)

Inria

Enhanced access to repository content

anHALytics

• Providing advanced reporting and analytical studies on scientific activities

Team: Achraf Azhar, Laurence Farhi, Patrice Lopez

http://traces1.saclay.inria.fr/anHALyticsv1/search/

Various scenarios:

• Research subjects of an institution, team or individual; evolution over time

• Collaboration patterns between institutions, countries, teams, individuals

• Domain overlaps and opportunities for multidisciplinary collaborations

• insights on new technical trends and emerging domains

HAL
archives-ouvertes.fr

+ add new facet

Q global warming   ⚙ ⌄   Disamb./Expand

publication_date ⌄

750 results - in 181 ms (server time)

93 96 97 99 03 04 05 06 07 08 09 10 11 12 13 14 15
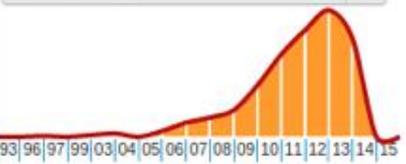
DD MM YYYY to DD MM YYYY ✔

subject-headers ⌄
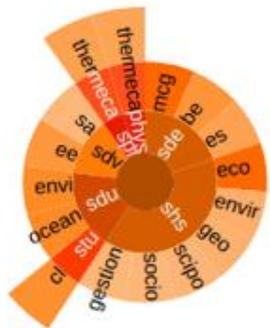
keywords ⌄

- climate change (54)
- changement climatique (20)
- adaptation (17)
- life cycle assessment (10)
- global warming (10)
- vulnerability (8)
- climate change mitigation (8)
- social sciences & humanities (7)
- sectoral policies (7)
- mitigation (7)

language ⌄

hal-00719481 - ENVIRONMENTAL EDUCATION for BEHAVIOUR **change: which actions should be targeted?**
E. Boyes, M. Stanisstreet - 20.7.2011
... with environmental problems that are believed to be both major and imminent, such as **global warming**; in such cases it is important...
... supporting more use of nuclear power, it was because even those students who believed it would reduce **global warming** would not...
... supporting more use of nuclear power, it was because even those students who believed it would reduce **global warming** would not...

hal-00682547 - Phenoloxidase and PEROXIDASE **activities in Sphagnum-dominated** PEATLAND **in a warming** CLIMATE
V. Jassey et al. - 1.3.2012
... ongoing **global warming**, these enzymes gain scientific concern in 37 terrestrial carbon reservoirs, such as peatlands (Fenner et al...
...Phenoloxidase and peroxidase activities in Sphagnum-dominated peatland in a **warming** climate Phenoloxidase and peroxidase...
...Phenoloxidase and peroxidase activities in Sphagnum-dominated peatland in a **warming** climate Phenoloxidase and peroxidase...

hal-00682547

JOURNAL ARTICLES

Vincent E. J. Jassey

Geneviève Chiapusio

Daniel Gilbert

Marie-Laure Toussaint

Philippe Binet

Abstract: PEATLANDS still suffer from the SCARCITY of available DATA about the CHARACTERIZATION and the response to CLIMATE FORCING of the main OXIDATIVE ENZYMES that occur over the seasons. In the present study, phenoloxidase and PEROXIDASE activities were examined in SPHAGNUM lawns along a narrow fen-bog GRADIENT under EXPERIMENTAL ELEVATED temperatures. We showed that PEROXIDASE activities from SPHAGNUM MOSSES were 1000-fold higher than those of phenoloxidases irrespective of seasons and SAMPLING areas. PEROXIDASE activities increased (+30%) with the rise of AIR temperatures (an AVERAGE of 1 °C), while warming did not alter phenoloxidase activities. These results suggest that the monitoring of PEROXIDASE activities in PEATLANDS may represent a suitable and forward indicator of the impact of CLIMATE WARMING on CARBON CYCLE in peatlands.

PEROXIDASE

Domains:
Biology

conf: 0.20
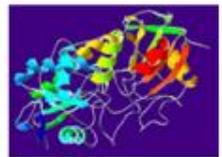
'''Peroxidases''' ([[EC number]] [http://www.chem.qmul.ac.uk/iubmb/enzyme/EC1/1.11.1.x]) are a large family of [[enzyme]]s that typically catalyze a reaction of the form:

Reference: W ⇌

hal-00965468 - TEMPERATURE OSCILLATION **coupled with** FUNGAL COMMUNITY **shifts can** MODULATE **warming effects on** LITTER DECOMPOSITION
C. Dang et al. - 31.1.2009
... feedbacks on climate (. The rise in mean temperatures anticipated by climate models dominates the current debate on **global**

ONE STEP BEYOND!

Repurposing publication repositories as digital libraries

Objectives

- Building up comprehensive scholarly collections that contribute to the preservation and dissemination of science

Various scenarios

- Personal collections, legacies
- Institutional heritage, patrimonies
- Community corpora
  - E.g.: IFIP Digital Library

Further decoupling publication repositories from traditional publishing

- E.g. re-publication of DL content

Challenges

Technical challenges

- Designing appropriate workflows (scanning, OCR, documentation)
- Working with heterogeneous formats; e.g. publishers' data
  - Importance of a highly structured pivot representation such as the TEI
- Cross-platform communication; e.g. conference management tools

Editorial challenges: challenging the traditional scope of publication repositories

- Variety of authors: no selection on provenance!
- A variety of objects: not just journal articles
  - Can we occasionally live with meta-data records?
- A variety of levels of scientificity: Nicéphore Niépce and shopping lists

Overlay journals with Episciences

Main functions of scholarly journals (Mabe, 2010)

Registration

Dissemination

Peer review

Archival record

Implementation in the Episciences model

*HAL: registration with precise affiliation information*

HAL: high visibility in search engines

Episciences: certification by editorial committees

HAL: long term archiving

Editorial workflow

Repository
publication

Repository: HAL,
arXiv, CWI…

Author

Submission to a
journal (editorial
board)

Comments,
interactions…

Reviewers

« Go »

épisciences.org

Ínría
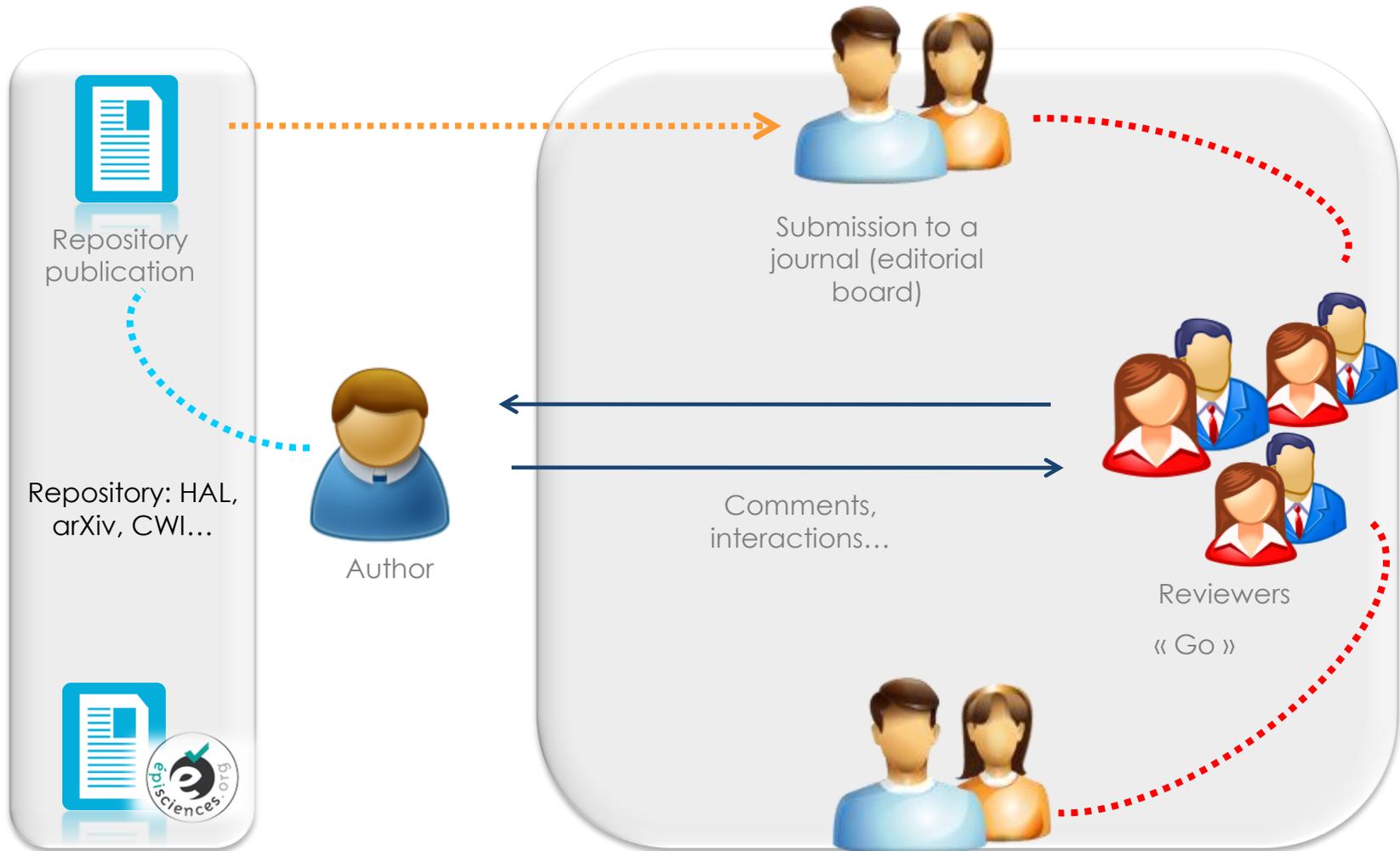
Episciences - issues

Acceptance of the post-publication peer-review model

- Manuscripts as a prior art reference
- Best coverage of research results (cf. clinical studies)

Sustainability

- Decoupling the sustainability of content from that of the certification platforms

Towards even more certification models

- Open peer review, author's initiative, community review

"Transitioning" from the existing journal eco-system

Homework

We should ban colours, not to say introducing new ones

We should stop fighting for open access
- institutions need real scientific information policies

The is no reason to preserve a system from another age that does not serve us (scholars) well

Installing a global vision of scientific communication as a research infrastructure
- Mutualising efforts in higher education and research

We should not forget data of course
- EU DARIAH agenda: data re-use charter

Thank you for your attention

Merci pour votre attention

"The current copyright regime hinders the exchange of knowledge and culture across borders."

https://juliareda.eu/copyright-evaluation-report-explained/
@Senficon