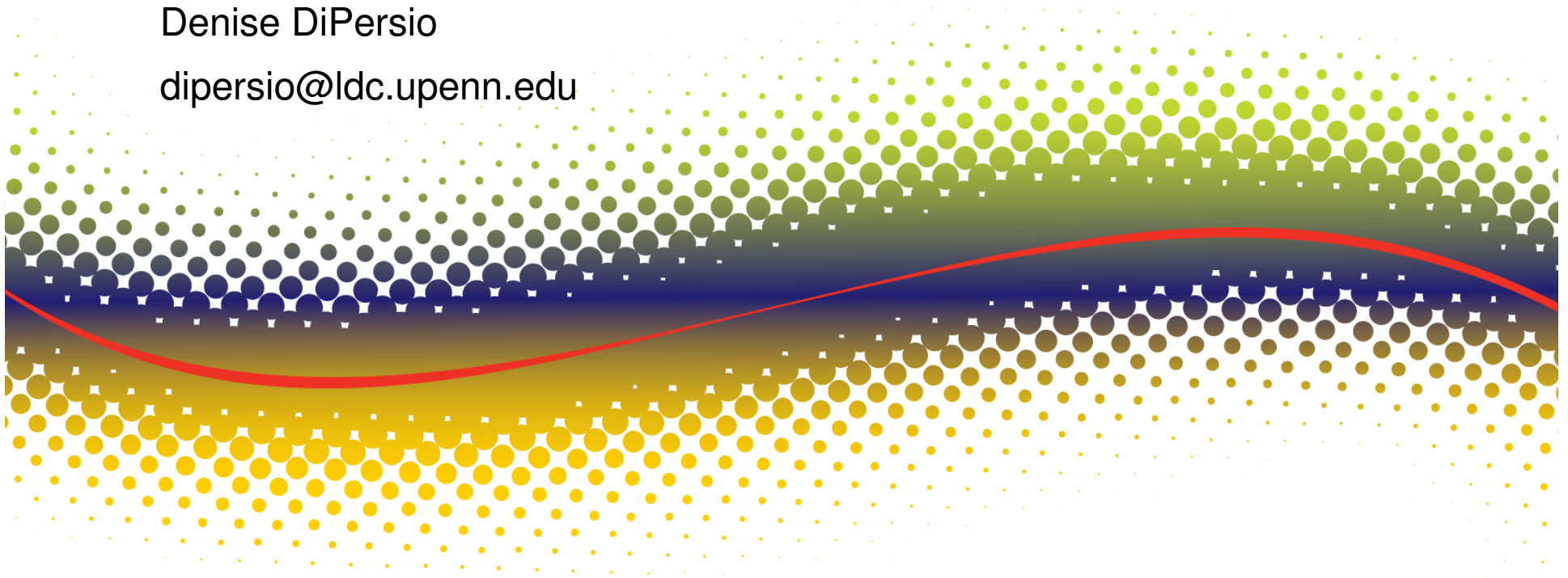




# Some Implications of US Initiatives for “Fair Research” and Open Access on the Development and Distribution of Language Resources

Denise DiPersio

[dipersio@ldc.upenn.edu](mailto:dipersio@ldc.upenn.edu)



- ◆ National Institutes of Health (NIH)
  - Requires electronic copies of peer-reviewed manuscripts to be deposited in National Library of Medicine Online Archive, PubMed Central, twelve months after publication in a peer-reviewed journal
  - Established in 2007 as part of US Congress appropriations bill; made permanent in 2009
  - Over 19 million citations; 3000 new manuscripts/month; 2 million visitors/day
  
- ◆ White House Office of Science and Technology Policy
  - Launched Public Access Policy Forum (2009): to solicit public views on access to publicly-funded research results
    - Research results include “data sets”, technical reports and peer-reviewed scholarly publications
    - Forum focusing on academic and scholarly journal articles
    - NIH model or other solutions
    - Comment period closed January 2010

- ◆ America COMPETES Act (2007; renewal pending 2010)
  - Initiatives to improve US competitiveness in mathematics, science and technology directed to various agencies, including NIST, National Science Foundation (NSF)
  - Section 7010 requires NSF to make available to the public timely and on NSF's website all final project reports and citations of published research documents resulting from NSF-funded research
    - NSF implementing requirement through FastLane Project Reports system; PIs must prepare public summary
    - Beginning October 2010, all NSF proposals must contain a data management plan (2 pages) for open data sharing under the proposed research (announced 05/10/2010)
    - NSF's longstanding policy that research results (data sets) be made publicly available at modest cost

- ◆ Bayh-Dole Act (1980)
  - University owns intellectual property it generates with government funds; government retains non-exclusive license
- ◆ Federal Acquisition Regulations (FARs)
  - Apply to US government contracts (v. grants)
  - Contractors require government's permission to claim copyright in materials "first produced" under the contract
  - Contractors grant government a non-exclusive license to reproduce, prepare derivative works, distribute, perform, display data not first produced under the contract (pre-existing material, perhaps copyrighted)
  - Alternate FAR clauses permit contractor to retain copyright in materials produced under contract with license to government and to designate pre-existing materials as "limited rights data"
- ◆ Federal purposes license
  - "Nonexclusive, worldwide, perpetual, irrevocable, royalty free and fully paid up license and right for the U.S. government to reproduce, publish or otherwise use the work, and to authorize others to do so, for Federal Government purposes"

- ◆ Fair Research in Copyright Act (2009)
  - Federal agencies cannot in funding agreements “impose or cause the imposition of any term of condition that requires the transfer or license to or for a Federal agency of any right provided under copyright law”
  - Described as the anti-NIH bill
  - Supported principally by publishing industry which claims NIH practice diminishes publishers’ copyrights in journal articles and undermines the peer review process
- ◆ Federal Research Public Access Act (2009, 2010)
  - Every department/agency with a research budget exceeding \$100 million must make research results available to the public within six months of their publication
  - Includes Departments of Commerce, Defense, Education and NSF
  - Public, private repositories can be used

## Fair Research in Copyright Act (H.R. 801)

- ◆ Applies to assertions of government rights in material developed under a funding agreement that “restrain or limit” copyright rights in an “extrinsic work”
  - Funding agreement = agreement between government and person receiving funds for research
  - Extrinsic work = a non-US government work “related” to funding agreement that includes “a meaningful added value” from nonparties to the funding agreement
- ◆ Opposed at hearings (2008) by NIH, Scholarly Publishing and Academic Resources Coalition (SPARC); supported by former US Register of Copyrights, American Physiological Association (American Association of Publishers)
- ◆ No action taken in 2008 pending further study; identical bill re-introduced in February 2009; referred to House Judiciary Committee, Subcommittee on Courts and Competition Policy

## ◆ Copyright issues

- Author as rights owner
- NIH negotiates agreement up front before manuscript submitted, preventing journal exclusivity

## ◆ Access issues

- NIH permits 12 month blackout period during which article only appears in journal before it is available on PubMed
- Journals claim that only 40% of an article's "use" occurs in first year
- Will libraries, institutions cancel subscriptions because of availability on PubMed? Lack of supporting evidence to date.
- PubMed as language resource data source

- ◆ Access
  - Research results available six months after publication (cf. NIH 12 month period)
  - Does not apply to progress reports, notes, preliminary data analyses, classified research, books (generate royalty, revenue), patentable discoveries, work not submitted to journals, work rejected by journals
- ◆ Repository-Agnostic, nonexclusive
  - Repositories can be public or private, provided they meet conditions of free access, interoperability and long-term preservation
  - Authors may deposit in multiple repositories
- ◆ Flexible approach to copyright, licensing
  - Agencies are to “make effective use of any law or guidance relating to the creation and preservation of a Government license that provides for the reproduction, publication, release, or other uses of a final manuscript for Federal purposes”
- ◆ Supported by universities, libraries; opposed by publishers
- ◆ Pending before Senate Committee on Homeland Security and Governmental Affairs (chair is a sponsor) and House Committee on Oversight and Government Reform
- ◆ Momentum for passage?



- ◆ Free access to public research in some form is mandated in many countries
  - Concern that intellectual property treaties are imperiled seems to have been put to rest
- ◆ Open access initiatives in academia – digital collections across disciplines
- ◆ LR community initiatives
  - Universal catalog – assumes ready access to resource superset
  - Language Commons – common license scheme for sharing and reuse
  - Building sustainable, interoperable language resources
    - The Human Language Project (Universal Corpus of the World's Languages)

- ◆ Current initiatives focus on peer-reviewed articles discussing research results
  - How can these principles be applied to research data, specifically, LRs?  
Or, should they be applied?
  - The NSF experience
- ◆ Role for data centers?
  - Building government infrastructure (e.g., PubMed) is costly
  - Harmonize licensing models
  - Costs -- distribution, storage, preservation
- ◆ Community input
  - European work well under way – CLARIN, FLaReNet
  - Tap into other projects, organizations