PROMOTING TRANSPARENCY IN SOCIAL SCIENCE RESEARCH

SESSION: OPEN DATA AND PROTECTION OF RESEARCH SUBJECTS

DECEMBER 12-13, 2013, UC BERKELEY

Mercè Crosas, Ph.D.
Twitter: @mercecrosas
Institute for Quantitative Social Science
Harvard University

IQSS

The Dataverse Network® Project

1. AN INTEGRATED PUBLISHING WORKFLOW FOR OPEN DATA

2. DATATAGS: SHARING EVEN SENSITIVE DATA WITH CONFIDENCE

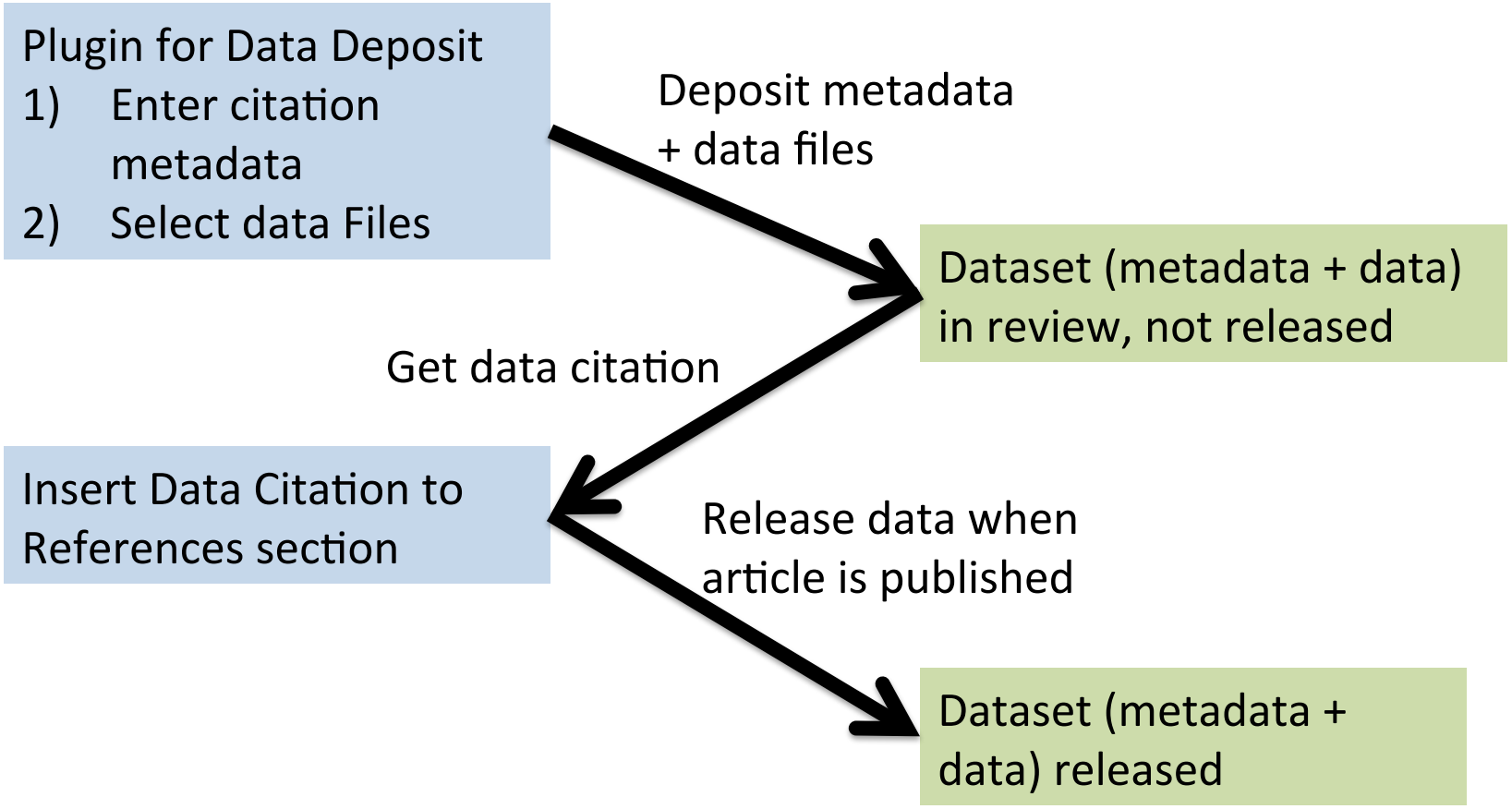
AN INTEGRATED PUBLISHING WORKFLOW FOR OPEN DATA

Enabling data deposit to a public repository as part of article's submission to a journal

http://projects.iq.harvard.edu/ojs-dvn

Journal's Publishing System

Article Submission Workflow



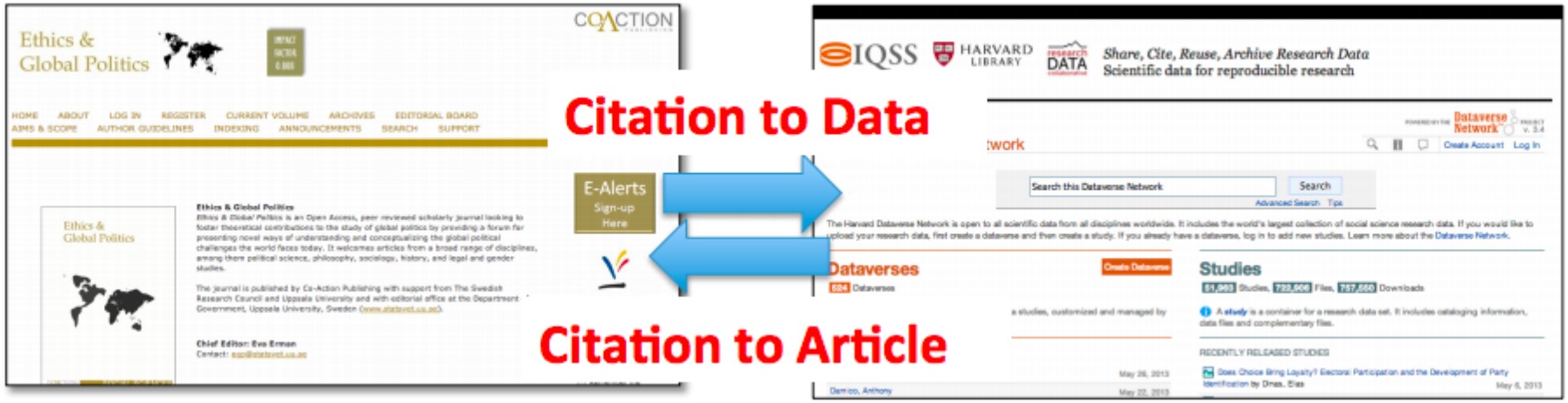
Open Data Repository

Data Sharing Workflow

Pilot Project: Integrate PKP's Open Journal System (OJS) with Dataverse

OJS Journal

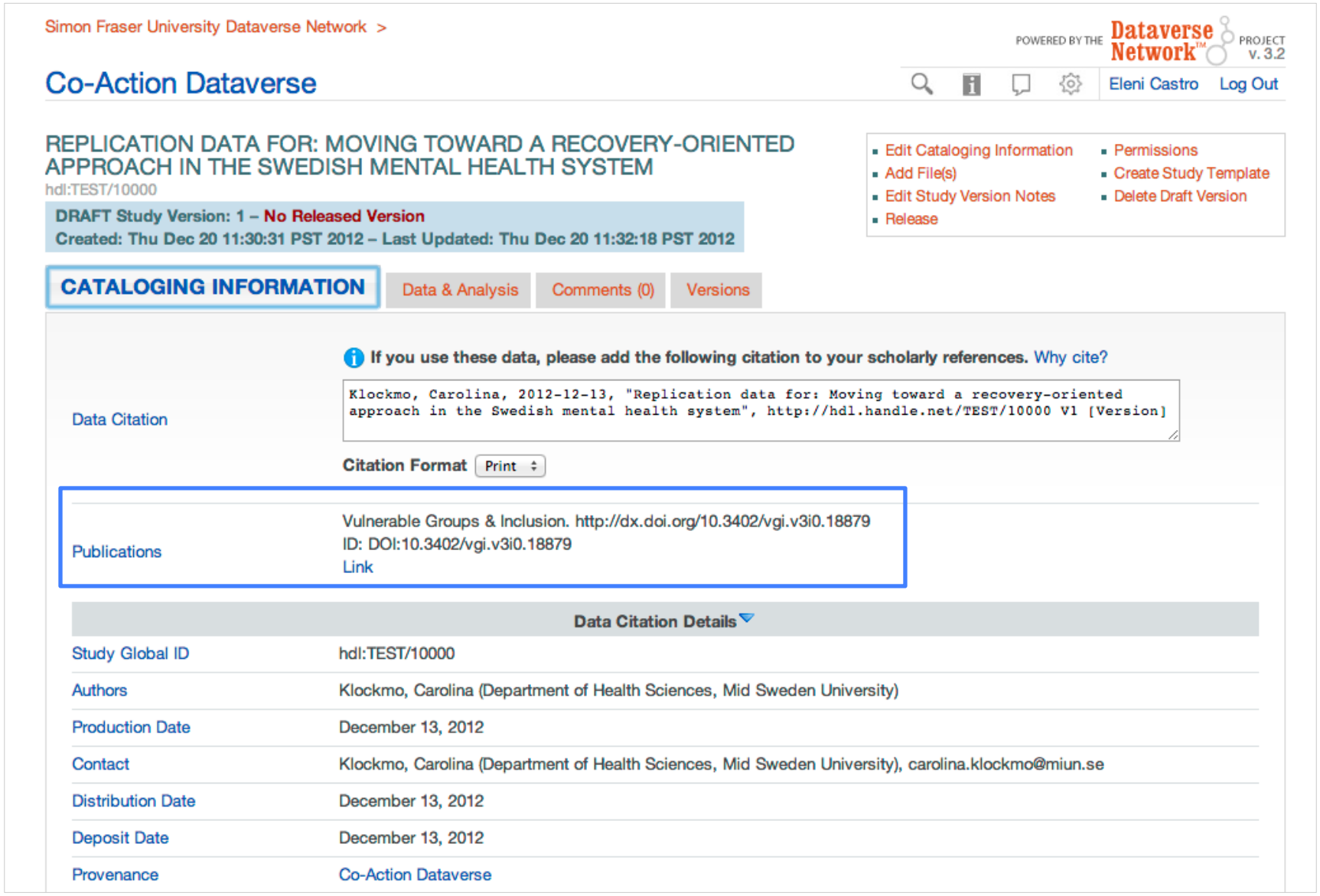
Harvard Dataverse Network



OJS plugin for:
Data + metadata + supporting files,
sent via SWORD API to the Dataverse

Published Article + Link to Data

Data in the Dataverse + Link to Article



Progress to-date

Phase I: Completed testing with 50+ journals from 6 Publishers.

Phase II: Extend testing to >100 journals (2014)

Scholars Portal
A Service of the Ontario Council of University Libraries

ubiquity press
open scholarship

COACTION PUBLISHING

UNIVERSITY OF PITTSBURGH
ULS
UNIVERSITY LIBRARY SYSTEM

UNM UNIVERSITY LIBRARIES

International Journal of Education Policy & Leadership

DATATAGS: SHARING EVEN SENSITIVE DATA WITH CONFIDENCE

A standardized framework for sharing data when data cannot be 100% open

Mercè Crosas, Latanya Sweeney, Michael Bar-Sinai

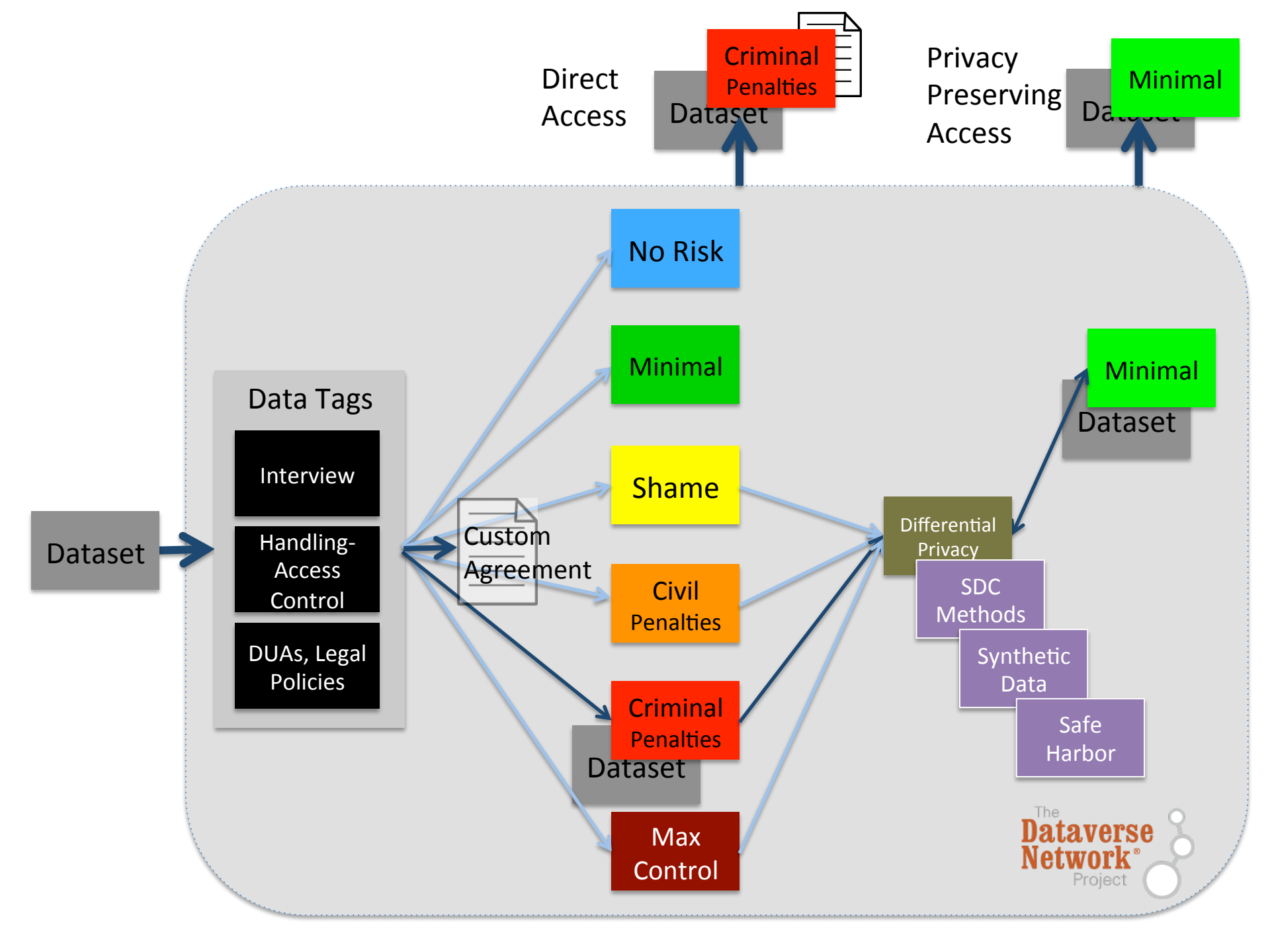
BERKMAN CENTER
AT HARVARD UNIVERSIT

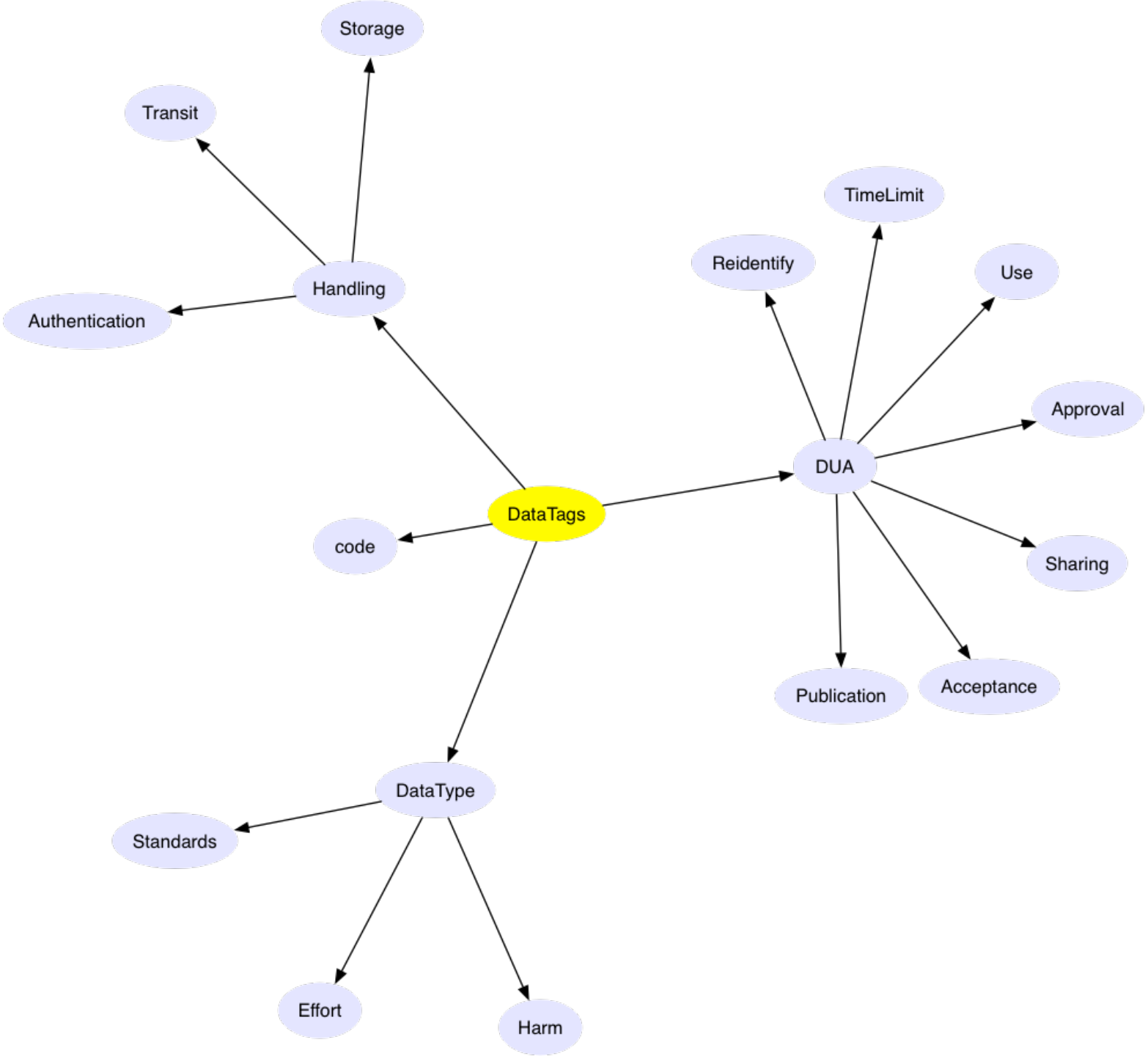
HARVARD
School of Engineering
and Applied Sciences

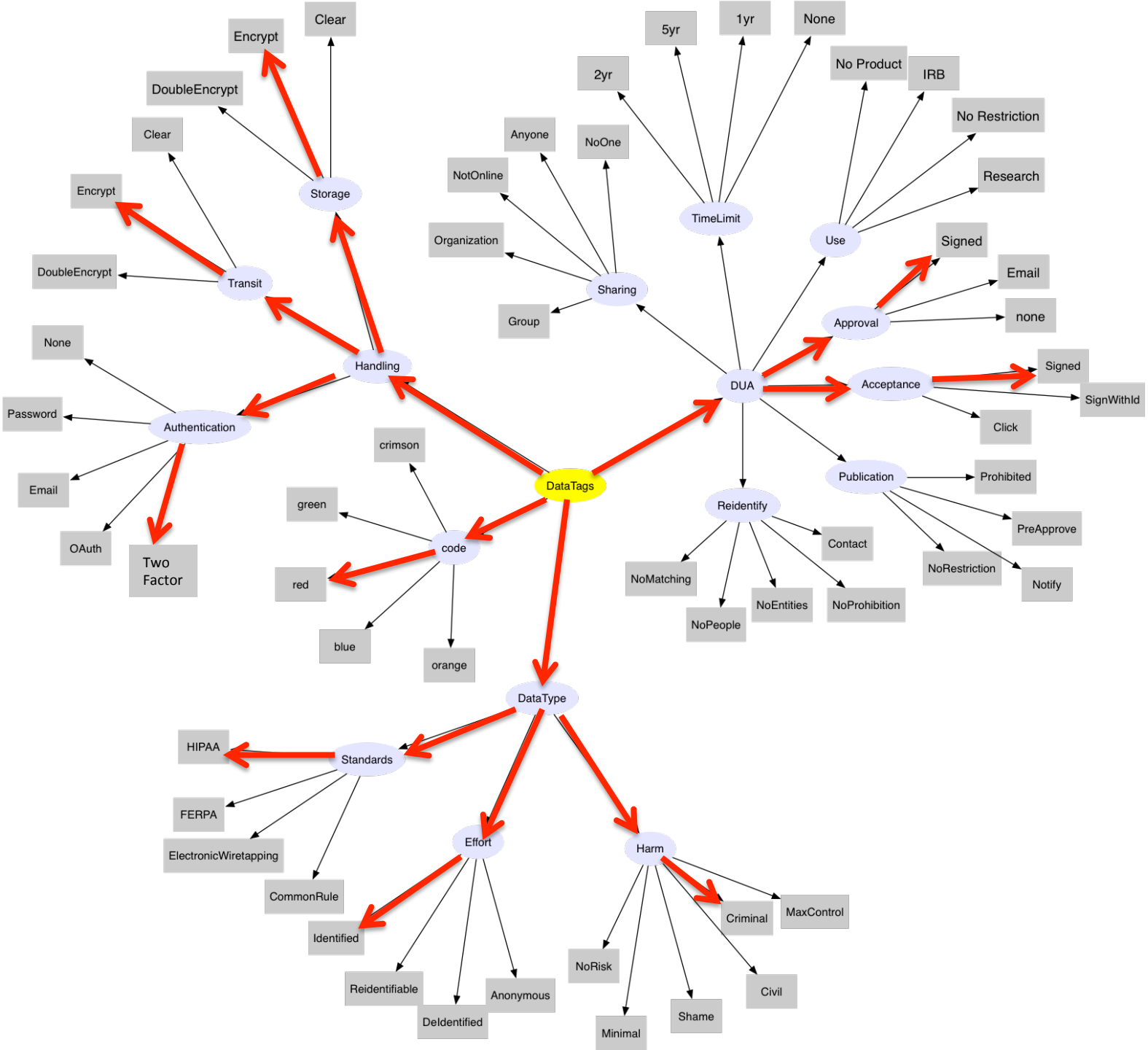
http://privacytools.seas.harvard.edu/

Tag Type	Description	Storage & Transit	Access
Blue	**Non-confidential** information, stored and shared freely.	Clear	Open
Green	**Not harmful** personal information, shared with some access control.	Clear	Email, OAuth verified registration
Yellow	**Potentially harmful** personal information, shared with loosely verified and/or approved recipients.	Clear & Encrypted	Password, Registered, Approval, click-through DUA
Orange	**Sensitive** personal information, shared with verified and/or approved recipients under agreement.	Encrypted	Password, Registered, Approval, signed DUA
Red	**Very sensitive** personal information, shared with strong verification of approved recipients under signed agreement.	Encrypted	Two-factor Auth, Registered, Approval, signed DUA
Crimson	**Maximum sensitive**, explicit permission for each transaction, strong verification of approved recipients under signed agreement.	Double Encrypted	Two-factor Auth, Registered, Approval, signed DUA

Extension of and compliant with Harvard Security Levels: http://security.harvard.edu/research-data-security-policy


Direct Access
Criminal Penalties
Dataset
Privacy Preserving Access
Minimal
Dataset
Dataset
Data Tags
Interview
Handling-Access Control
DUAs, Legal Policies
Custom Agreement
No Risk
Minimal
Shame
Civil Penalties
Criminal Penalties
Dataset
Max Control
Differential Privacy
SDC Methods
Synthetic Data
Safe Harbor
Minimal
Dataset
The Dataverse Network® Project


Storage
Transit
Handling
Authentication
TimeLimit
Reidentify
Use
Approval
DUA
DataTags
code
Sharing
Publication
Acceptance
DataType
Standards
Effort
Harm


DataTags
Storage
Clear
Encrypt
DoubleEncrypt
Clear
Encrypt
Transit
DoubleEncrypt
Handling
None
Password
Authentication
Email
OAuth
Two Factor
DUA
TimeLimit
5yr
1yr
None
2yr
Anyone
NoOne
NotOnline
Organization
Sharing
Group
Use
No Product
IRB
No Restriction
Research
Approval
Signed
Email
none
Acceptance
Signed
SignWithId
Click
Publication
Prohibited
PreApprove
NoRestriction
Notify
Reidentify
Contact
NoMatching
NoPeople
NoEntities
NoProhibition
code
crimson
green
red
blue
orange
DataType
Standards
HIPAA
FERPA
ElectronicWiretapping
CommonRule
Effort
Identified
Reidentifiable
DeIdentified
Anonymous
Harm
NoRisk
Minimal
Shame
Criminal
Civil
MaxControl

Privacy Laws

- 2187 privacy related laws exist in the United States, which can be grouped in ~ 30 types:

  Explicit consent, medicals records, student records, arrest and conviction records, bank and financial records, cable television, computer crime, credit reporting, criminal justice, electronic surveillance, employment records, government information on individuals, identity theft, insurance records (including use of Genetic information), library records, mailing lists, special medical records (including HIV testing), non-electronic surveillance, polygraphing in employment, privacy statutes/state constitutions, privileged communications, social security numbers, tax records, telephone services, testing in employment, tracking technologies, voter records.

Acknowledgement: Robert Smith, Latanya Sweeney

Phase I Implementation, Education and Outreach

- DataTags.org web application, initially supporting HIPAA regulations:
  - Explicit Consent
  - Medical Records
- Integrate with Secure **Dataverse**
- Alpha, Beta testing with hundreds of Social Scientists
- YouTube videos on how to use it

On a parallel track:

- HIPAA certification
- Pilot at Harvard Medical School and School of Public Health

DATA TAGS DEMO

http://datatags.org

Mercè Crosas, Latanya Sweeney, Michael Bar-Sinai