Data Repository Assessment & Certification at ICPSR: Experiences and Lessons Learned

Jared Lyle
NIH Trustworthy Data Repositories workshop
Rockville, Maryland
April 8, 2019

Acknowledgements

- Mary Vardigan
- Nancy McGovern

Outline

- Overview of ICPSR

- Why assessment is important

- ICPSR's experience with assessment, including effort and resources needed

- Benefits from assessment

http://www.icpsr.umich.edu

ICPSR

Philip Converse, Warren Miller, and Angus Campbell

- Established 1962
- Originally 22 Members, now consortium of 776 world-wide
- Originally Political Science, now all social and behavioral sciences

ICPSR

- Current holdings
    - 10,000+ studies, quarter million files
    - 1500+ are *restricted studies*, almost always to protect confidentiality
    - Bibliography of Data-related Literature with 80,000 citations
- Approximately 60,000 active MyData ("shopping cart") accounts
- Thematic collections of data about addiction and HIV, aging, arts and culture, child care and early education, criminal justice, demography, health and medical care, and minorities

Funders

- Administration for Children and Families, DHHS
- Alfred P. Sloan Foundation
- American Educational Research Association
- American Psychological Association
- Andrew W. Mellon Foundation
- Annie E. Casey Foundation
- Bill & Melinda Gates Foundation
- Bureau of Justice Statistics, DOJ
- Centers for Disease Control and Prevention, Department of Health and Human Services
- Federal Highway Administration
- Institute of Education Sciences, DOE
- Laura and John Arnold Foundation
- National Center for Health Statistics
- National Endowment for the Arts
- National Human Genome Research Institute, NIH

- National Institute of Allergy and Infectious Diseases, NIH
- National Institute of Child Health and Human Development, NIH
- National Institute of Justice, DOJ
- National Institute on Aging, NIH
- National Institute on Drug Abuse, NIH
- National Science Foundation
- NSD-Norwegian Centre for Research Data, Ltd.
- Office of Juvenile Justice and Delinquency Prevention, DOJ
- Patient-Centered Outcomes Research Institute
- Robert Wood Johnson Foundation
- Spencer Foundation
- U.S. Agency for International Development

Midlife in the United States (MIDUS 1), 1995-1996 (ICPSR 2760)

Version Date: Mar 25, 2019 ❓ | Cite this study | Share this page

Principal Investigator(s): ❓

Orville Gilbert Brim, John D. and Catherine T. MacArthur Foundation. Research Network on Successful Midlife Development; Paul B. Baltes, Max-Planck-Institut für Bildungsforschung.; Larry L. Bumpass, University of Wisconsin; Paul D. Cleary, Harvard Medical School; David L. Featherman, University of Michigan; William R. Hazzard, Wake Forest University; Ronald C. Kessler, Harvard Medical School; Margie E. Lachman, Brandeis University; Hazel Rose Markus, Stanford University; Michael G. Marmot, University College, London. Medical School; Alice S. Rossi, University of Massachusetts at Amherst; Carol D. Ryff, University of Wisconsin; Richard A. Shweder, University of Chicago

Series:

- Midlife in the United States (MIDUS) Series

https://doi.org/10.3886/ICPSR02760.v15

Version V15 (see more versions)

Download ▾	📈 Analyze Online

At A Glance	Data & Documentation	Variables	Data-related Publications	Export Metadata

▾ **Project Description**

Alternate Title ❓

MIDUS 1

Summary ❓

The Midlife in the United States (MIDUS) is a collaborative, interdisciplinary investigation of patterns, predictors, and consequences of midlife development in the areas of physical health, psychological well-being, and social responsibility. A description of the study and findings from it are available at http://www.midus.wisc.edu ↗. The MIDUS 1 data collection is comprised of four parts.

Part 1, Main, Siblings, and Twin Data, contains responses from the main survey of 7,108 respondents. Respondents were asked to provide extensive information on their physical and mental health throughout their adult lives, and to assess the ways in which their lifestyles, including relationships and work-related demands, contributed to the conditions experienced. Those queried were asked to describe their histories of physical ailments, including heart-related conditions and cancer, as well as the treatment and/or lifestyle changes they went through as a result. A series of questions addressed alcohol, tobacco, and illegal drug use, and focused on history of use, regularity of use, attempts to quit, and how the use of those substances affected respondents' physical and mental well-being. Additional questions addressed respondents' sense of control over their health, their awareness of changes in their medical conditions, commitment to regular exercise and a healthy diet, experience with menopause, the decision-making process used to deal with health concerns, experiences with nontraditional remedies or therapies, and history of attending support groups. Respondents were asked to compare

Search Variables

[_____] GO

Variables

A1SA9U Alcohol/drug problem ever (12 months)
A1SA9V Migraine headaches ever (12 months)
A1SA9W Chronic sleep problems ever (12 months)
A1SA9X Diabetes/high blood sugar ever (12 months)
A1SA9Y Neurological disorder ever (12 months)
A1SA9Z Stroke ever (12 months)
A1SA9AA Ulcer ever (12 months)
A1SA9BB Hernia ever (12 months)
A1SA9CC Piles/hemorrhoids ever (12 months)
A1SCHRON Number of chronic conditions (12 months)
A1SCHROX Having any chronic conditions (12 months)
A1SA10A Prescription medicine for hypertension ever (30 days)
A1SA10B Prescription medicine for diabetes ever (30 days)
A1SA10C Prescription medicine for cholesterol ever (30 days)
A1SA10D Prescription medicine for heart condition ever (30 days)
A1SA10E Prescription medicine for lung problems ever (30 days)
A1SA10F Prescription medicine for ulcer ever (30 days)

A1SA9U: Alcohol/drug problem ever (12 months)

Question: In the past 12 months, have you experienced or been treated for any of the following? Alcohol or drug problems

Value	Label	Unweighted Frequency	%
1	Yes	161	2.3%
2	No	6137	86.3%
	Missing Values		
-1	Respondent does not have SAQ data	783	11.0%
8	Refused	27	0.4%
	Total	7108	100%

Based upon 6298 valid cases out of 7108 total cases.

Summary Statistics

- mode: 2.00
- median: 2.00
- minimum: 1.00
- maximum: 2.00
- mean: 1.97
- standard deviation: 0.16

Variable Type: numeric

(Range of) Missing Values: -1, 8

Notes

Source: This variable was taken from Midlife in the United States (MIDUS 1), 1995-1996.

Copyright: ICPSR has an FAQ on copyright and survey instruments.

Disclaimer: The frequencies for this variable *may not be weighted.* They are purely descriptive and may not be representative of the study population. Please use with caution and consult the study documentation.

```
 8010030000200000000000000000070000100001000000000000000000110000229974  2 12 1 177977997 25
 900000000000000003000000000700000000000000001000000000000000100112273  2   8 1 177977997 38
1000000000300001000100000007000000001000000000000000001000011  756 2 17 1 577977997 12
1100000000300000000000000007000000001000000000000000001000011  750 3   4141377977997  1
12000000000000003000000700000000000010000000000000000100111157  3  75131377977997  1
1300000
1400000  1    5 1120 120865002372819912 01  5082006
1500000  2    5 1120 120865002372819912 011  198111177 1 111 191399999999309 23911222411240111224212999999
1600000  3    5 1120 120865002372819912 1  121140401111122424112 110200199999 3 129992999299929992999299992
1700000  3    5 1120 120865002372819912 2 22 23939112222411112 111200199999 3 129999999999992999299929999
1830000  1    6 11208942977521295319912 001  5082006
1900000  2    6 11208942977521295319912 011  198112222 1 111 19989999999993091162112212112621122121129999
2011000  3    6 112
2110100  3    6 112
2200000  3    6 112
2300100  1    7 112
2400000  2    7 112
2500000  3    7 112
2600000  4    7 112
2700000  3    7 112
2800000  1    8 112
2900000  2    8 112
3000000  3    8 112
3102000  3    8 112
        3    8 112
        1    9 112
        2    9 112
        3    9 112
        1   10 112
        2   10 112
        3   10 112
        1   11 112
        2   11 112
        3   11 112
```

Count	Card	Height	Gender	Age	Eye color	Hand	Siblings	Coffees	TV time	Die 1	Die 2	Satisfaction
1	1	177.8	male	18	green	right	4	3	0.0	3	2	4
2	2	170.2	female	23	blue	left	3	3	0.0	7	5	4
3	3	185.4	male	17	blue	right	1	3	1.0	3	8	4
4	4	185.4	male	22	brown	right	2	1	0.0	3	2	5
5	5	160.0	female	20	brown	right	2	0	1.0	3	2	4
6	6	175.3	male	20	brown	right	1	1	2.0	3	2	3
7	7	177.8	male	23	brown	right	2	1	1.0	6	2	3
8	8	185.4	male	23	green	right	3	0	2.0	1	5	4
9	9	160.2	female	20	blue	right	2	0	2.5	2	3	4
10	10	182.8	male	21	brown	right	2	0	0.0	4	3	5
11	12	162.6	female	21	blue	right	2	0	1.0	5	6	5
12	13	172.7	male	19	green	right	2	0	4.0	5	6	4
13	16	172.7	male	19	green	right	1	0	1.0	3	4	3
14	19	165.1	female	25	brown	right	1	0	1.0	2	3	3
15	20	162.6	female	20	blue	right	4	2	1.0	4	2	3
16	21	175.3	male	17	brown	right	3	0	1.5	6	2	3
17	22	180.3	male	17	black	left	3	0	0.5	1	1	2
18	23	185.4	male	20	brown	right	2	2	0.0	5	6	4
19	25	188.0	male	20	blue	right	1	3	0.0	6	6	4
20	26	170.2	male	20	blue	right	1	0	2.0	6	5	4
21	27	154.9	female	18	brown	right	1	0	2.0	4	3	4
22	28	190.5	male	19	blue	right	2	0	0.5	3	2	4
23	29	178.0	male	25	blue	right	2	0	2.0	5	2	5
24	30	170.2	female	20	brown	right	2	0	0.0	2	4	4
25	31	177.8	male	19	blue	right	1	0	1.0	2	3	3
26	33	160.0	female	19	brown	right	0	1	1.0	3	3	3
27	34	175.3	male	26	brown	right	1	0	0.0	3	5	4

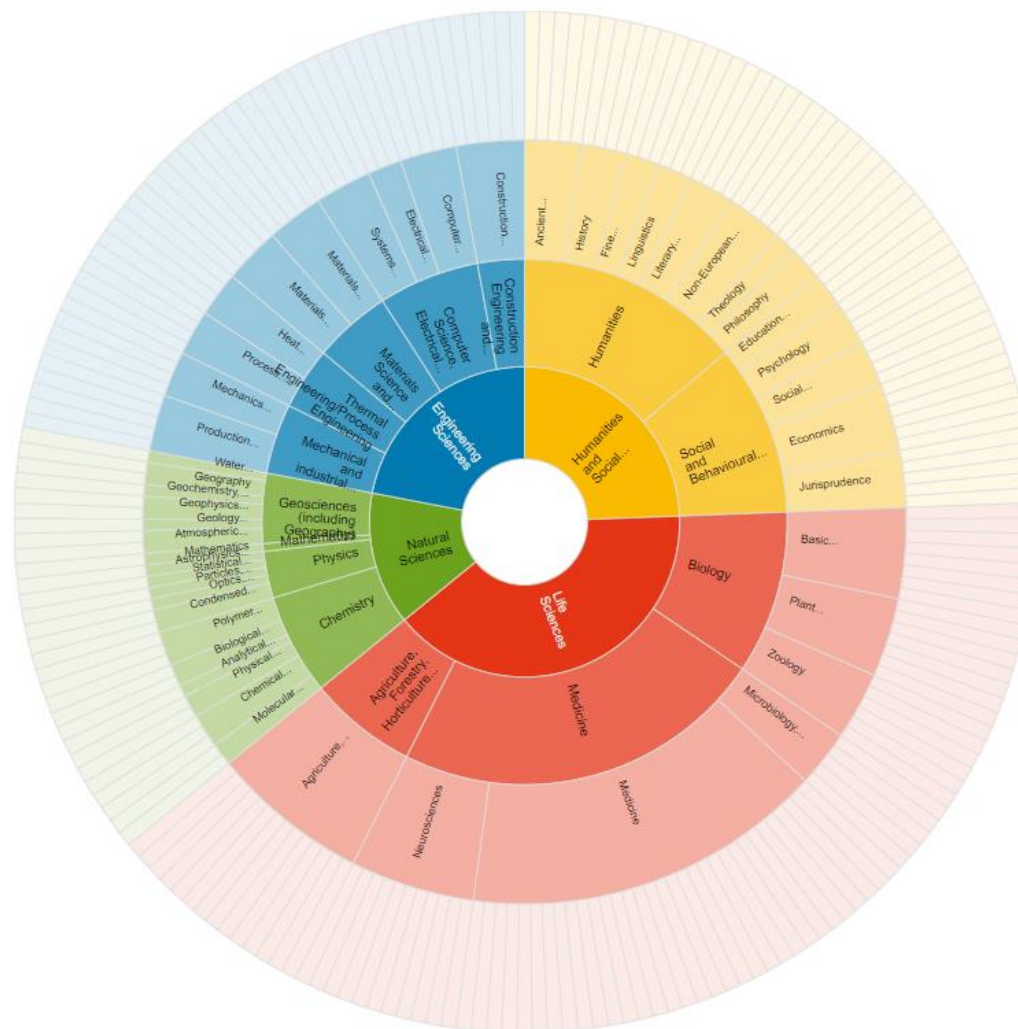# Why Assessment is Important

- Provide transparent view into the repository
- Improve processes and procedures
- Measure against a community standard
- Show the benefits of domain repositories

- Promote trust by funding agencies, data producers, and data users that data will be available for the long term

Dillo, I., & de Leeuw, L. (2018). CoreTrustSeal. Communications of the Association of Austrian Librarians, 71(1), 162-170. https://doi.org/10.31263/voebm.v71i1.1981

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Search...   🔍 Search

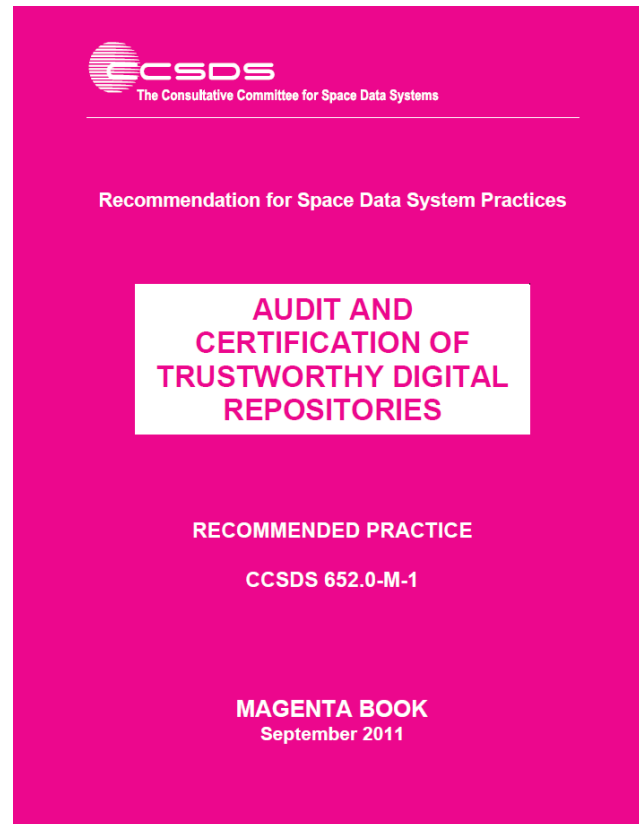# Forever!

Guaranteed!

We promise!

"Claims of trustworthiness are easy to make but are thus far difficult to justify or objectively prove."

Comments (51)

Hazards of the Cloud: Data-Storage Service's Crash Sets Back Researchers

Dedoose, a cloud-based application for managing research data, suffered a "devastating" technical failure last week that caused academics across the country to lose large amounts of research work, some of which may be gone for good.

SocioCultural Research Consultants, the company that sells Dedoose, is still scrambling to recover as much of its customers' work as possible, and has said in a blog post that "the vast majority" of research data on its platform were not affected.

The crash nonetheless has dealt frustrating setbacks to a number of researchers, highlighting the risks of entrusting data to third-party stewards.

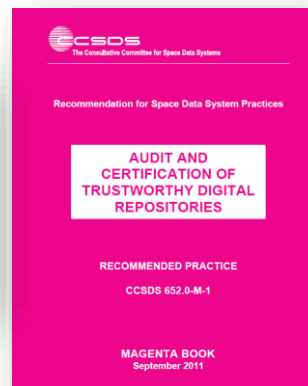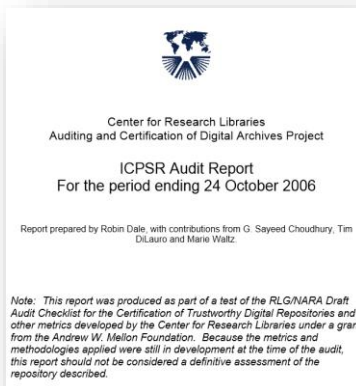http://chronicle.com/blogs/wiredcampus/hazards-of-the-cloud-data-storage-services-crash-sets-back-researchers/52571

If we want to be able to share data, we need to store them in a trustworthy data repository. Data created and used by scientists should be managed, curated, and archived in such a way to preserve the initial investment in collecting them. Researchers must be certain that data held in archives remain useful and meaningful into the future.

"An Introduction to the Core Trustworthy Data Repositories Requirements"

ICPSR Assessment Experience

2005-2006   CRL test audit (TRAC checklist)

2010-2012   TRAC/ISO 16363 self-assessment

2009-2010   Data Seal of Approval certification

2013   Data Seal of Approval (update)

2013   World Data System certification

2018-2019   CoreTrustSeal

CRL Test Audit, 2005-2006

- Test methodology based on RLG-NARA Checklist for the Certification of Trusted Digital Repositories

- Assessment performed by an external agency (CRL)

- Precursor to current TRAC audit/certification

- ICPSR Test Audit Report: http://www.crl.edu/sites/default/files/attachments/pages/ICPSR_final.pdf

Effort and Resources Required

- Completion of Audit Checklist
- Gathering of large amounts of data about the organization – staffing, finances, digital assets, process, technology, security, redundancy, etc.
- Weeks of staff time to do the above
- Hosting of audit group for two and a half days with interviews and meetings
- Remediation of problems discovered

Findings

Positive review overall:

*Taken as a whole, ICPSR appears to provide **responsible stewardship** of the valuable research resources in its custody. **Depositors of data to the ICPSR data archives and users of those archives can be confident** about the state of its operation, and the processes, procedures, technologies, and technical infrastructure employed by the organization.*

Findings

Positive review overall, but…

- Succession and disaster plans needed
- Funding uncertainty (grants)
- Acquisition of preservation rights from depositors
- Need for more process and procedural documentation related to preservation
- Machine-room issues noted

Changes Made

- Hired a Digital Preservation Officer
- Created policies, including Digital Preservation Policy Framework, Access Policy Framework, and Disaster Plan
- Changed deposit process to be explicit about ICPSR's right to preserve content
- Continued to diversify funding (ongoing)
- Made changes to machine room

TRAC self-assessment, 2010-2012

- TRAC/ISO most rigorous method – 80+ requirements (100+ in ISO)
- OAIS orientation

Procedures Followed

- Parceled out the 80+ TRAC requirements to committees across the organization

- Set up system for reporting evidence

- Gathered evidence demonstrating compliance for each guideline; rated compliance on scale

- Digital Preservation Officer and Director of Curation Services reviewing evidence

Example TRAC/ISO Requirements

- *Documented process for testing understandability of the information content*

- *Dissemination of authentic copies of the original or objects traceable to originals*

Effort and Resources Required

- Time of many individuals across the organization
- Time for high-level review and summarization
- Time/technology most likely required to address areas for improvement

DSA Self-Assessment, 2009-2010

Data Seal of Approval

- Started by DANS in 2009
- The objectives of the DSA are to "safeguard data, to ensure high quality and to guide reliable management of data for the future without requiring the implementation of new standards, regulations, or high costs."

http://www.datasealofapproval.org/en/information/about/

Data Seal of Approval

- 16 guidelines – 3 target the data producer, 3 the data consumer, and 10 the repository

- Example guideline: *(7) The data repository has a plan for long-term preservation of its digital assets.*

- Self-assessments are done online with ratings and then peer-reviewed by a DSA Board member

Procedures Followed

- Digital Preservation Officer and Director of Collection Delivery conducted self-assessment, assembled evidence, completed application
- Provided a URL for each guideline

Effort and Resources Required

- Mainly time of the Digital Preservation Officer and Director of Collection Delivery
- Would estimate two days at most
- Less time required to recertify every two years

Self-Assessment Ratings

- Using the manual and guiding questions: Rated ICPSR as having achieved 4 stars for all but Guideline 13, which addresses full OAIS compliance

Findings and Changes Made

- Recognized need to make policies more public – e.g., static and linkable Terms of Use (previously only dynamic)

- Reinforced work on succession planning – now integrated into Data-PASS partnership agreement

- Underscored need to comply with OAIS – building a new system based on it

DSA Self-Assessment, 2014-2015

https://assessment.datasealofapproval.org/assessment_114/seal/pdf/
http://hdl.handle.net/2027.42/144319

World Data System Certification, June 2013

- WDS is effort of the International Council of Science (ICSU)

- Started in natural sciences -- similar to Data Seal of Approval

- Membership and certification mechanisms

World Data System Certification, June 2013

- 20+ criteria (guidelines)
- Example criterion: *The facility ensures integrity and authenticity of data sets during ingest, archival storage, data quality assessment and analysis, product generation, access, and delivery*

Effort and Resources Required

- Time of one individual – around two days
- Five-stage process: Organization expresses interest; demonstrates its capabilities; if necessary, an on-site review may occur; accreditation; review every 3-5 years

Findings

- Permitted comparison of WDS and DSA content and procedures

- Resulted in WDS-DSA Working Group under the umbrella of the RDA Certification IG

- WG assessed commonalities and potential to combine efforts, which resulted in the CoreTrustSeal Data Repository certification

CoreTrustSeal, 2018-2019

CoreTrustSeal

- Developed by the DSA-WDS Partnership Working Group on Repository Audit and Certification, a Working Group of the Research Data Alliance

- Merging of the Data Seal of Approval certification and the World Data System certification

- 16 criteria (guidelines)

Requirements

- 16 criteria (guidelines):
    - Organizational Infrastructure (6)
    - Digital Object Management (8)
    - Technology (2)

Example of Evidence – R5

- Guideline Text: R5. The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission

Example of Evidence – R5

Guidance: The range and depth of expertise of both the organization and its staff, including any relevant affiliations (e.g., national or international bodies), is appropriate to the mission.

ICPSR Response: R5 (one part)

A 12-person Council whose members are elected by the ICPSR membership provides guidance and oversight to ICPSR. Members serve four-year terms, and six new members are elected every two years. The Council acts on administrative, budgetary, and organizational issues on behalf of all the members of ICPSR. [6]

ICPSR's staff of over 100 perform a variety of functions to support ICPSR's archival and training missions. The staff include data curators and managers, librarians, Web developers, communications specialists, user support specialists, administrative staff, and a small team of researchers, as well as software developers, programmers, system administrators, and desktop support specialists. Staff have expertise in digital archiving, data preservation, usability testing, Section 508 review for ADA Section 8 compliance, DOI registration, web traffic analytics, search engine optimization, storage and dissemination of sensitive data, restricted-use data agreements, and researcher credentialing. All staff are required to complete ongoing training related to data security and disclosure risk. [7]

ICPSR Response: R5 (references)

References:

[1] ICPSR Web site, About the Organization: https://www.icpsr.umich.edu/icpsrweb/content/about/index.html (accessed 2018-10-04)

[2] ICPSR 2016-2017 Annual Report, Financial Reports: https://www.icpsr.umich.edu/files/ICPSR/about/annualreport/2016-2017.pdf (accessed 2018-11-08)

[3] ICPSR Web site, Thematic Data Collections: https://www.icpsr.umich.edu/icpsrweb/content/about/thematic-collections.html (accessed 2018-10-04)

[4] ICPSR Web site, List of Member Institutions and Subscribers: https://www.icpsr.umich.edu/icpsrweb/membership/administration/institutions (accessed 2018-11-06)

…

Effort and Resources Required

- 3-5 days of time by the Director of Metadata and Preservation
- Less time required to certify every 3 years

Dear Jared Lyle,

Thank you! We have successfully received your application on behalf of Inter-university Consortium for Political and Social Research to be certified under the CoreTrustSeal. Your application has now been locked, and further editing of its contents is not possible at this time. A URL to access the application for viewing or printing can be found here: https://amt.coretrustseal.org/inter-university-consortium-for-political-and-social-research/1

This application has the status of Application Submitted. Please note that if this the first submission of the current application, the review process will begin only after payment has been confirmed of the 1000 EUR administration fee by Inter-university Consortium for Political and Social Research. The legal entity managed by the CoreTrustSeal Secretariat will notify you as to how this fee should be paid. Upon receipt, the application will be sent for review, and we will contact you again with the decision of the CoreTrustSeal Standards and Certification Board, once available. We kindly ask for your patience during this time.

You added the follow additional comments to your application:

Best regards,
CoreTrustSeal Secretariat
(On behalf of the CoreTrustSeal Standards and Certification Board)

CoreTrustSeal Secretariat
The Hague, Netherlands | Tokyo, Japan
+31 6 2386 3243 | +81 4 2327 6395
info@coretrustseal.org

Dear Jared,

On behalf of the CoreTrustSeal Standards and Certification Board, I would like to ask you the following details with respect to the CoreTrustSeal application for Inter-university Consortium for Political and Social Research.

- ·   name of the organization to which the invoice can be addressed
- ·   organization address to which the invoice can be addressed
- ·   email address that can be used for sending the invoice

Upon receipt, the CoreTrustSeal Secretariat will prepare the invoice for the administrative fee, see https://www.coretrustseal.org/apply/administrative-fee/ .

Thank you in advance.

Should you have any questions, do not hesitate to contact me.

Best wishes,
Lisa de Leeuw
(On behalf of the CoreTrustSeal Standards and Certification Board)

finance@CoreTrustSeal.org | www.CoreTrustSeal.org

Dear Jared,

This is to confirm that Foundation CoreTrustSeal received the administrative fee from ICPSR for 1 CoreTrustSeal Data Repository Certification in good order.

The CoreTrustSeal Secretariat will assign reviewers shortly that will be given a deadline of maximum two months to provide their thoughts to the CoreTrustSeal Board, which will then make a decision at its subsequent meeting. Afterwards, you will be informed on their decision and/or feedback.

Best wishes,
Lisa de Leeuw
(On behalf of the CoreTrustSeal Standards and Certification Board)

finance@CoreTrustSeal.org | www.CoreTrustSeal.org

Findings and Changes Made

- In progress -- CoreTrustSeal Secretariat still completing review
- Some fine tuning:
    - Making older versions of collections available
    - File-level persistent identifiers and citations

- Uncertainty about the level of detail required

Comparison of Assessments – Effort and Resources

- Test audit was the most labor- and time-intensive
- TRAC self-assessment involved the time of more people
- CoreTrustSeal (Data Seal of Approval and World Data System) certification least costly

Comparison of Assessments – Benefits

- What did we learn and did the results justify the work required?
  - Test audit was first experience – resulted in greatest number of changes, greatest increase in awareness
  - Fewer changes made as a result of CoreTrustSeal (DSA and WDS); also not as detailed
  - TRAC assessment has surfaced additional issues to address

Benefits continued

- Difficult to quantify
  - Trust of stakeholders
  - Transparency
  - Teaching opportunity for new staff
  - Improvements in processes and procedures
  - Use of community standards and alignment across domains

- Leadership dimension also important

Thank you!

lyle@umich.edu