

The Dataverse Network: An Infrastructure for Data Sharing

Gary King
Harvard University

February 5, 2008

- Gary King, **An Introduction to the Dataverse Network as an Infrastructure for Data Sharing**, *Sociological Methods and Research*, 32, 2 (November, 2007): 173–199.

- Gary King, **An Introduction to the Dataverse Network as an Infrastructure for Data Sharing**, *Sociological Methods and Research*, 32, 2 (November, 2007): 173–199.
- Micah Altman and Gary King. **A Proposed Standard for the Scholarly Citation of Quantitative Data**, *D-Lib Magazine*, 13, 3/4 (March/April).

- Gary King, **An Introduction to the Dataverse Network as an Infrastructure for Data Sharing**, *Sociological Methods and Research*, 32, 2 (November, 2007): 173–199.
- Micah Altman and Gary King. **A Proposed Standard for the Scholarly Citation of Quantitative Data**, *D-Lib Magazine*, 13, 3/4 (March/April).
- Dataverse Network project: <http://TheData.org>

Infrastructure for Quantitative Data

- Accessibility:

- Accessibility:
 - Most large data sets: in public archives

- Accessibility:
 - Most large data sets: in public archives
 - Most data in published articles: not accessible, results not replicable without the original author

Infrastructure for Quantitative Data

- Accessibility:
 - Most large data sets: in public archives
 - Most data in published articles: not accessible, results not replicable without the original author
 - Most data sets from NSF & NIH grants: not publicly available

Infrastructure for Quantitative Data

- Accessibility:
 - Most large data sets: in public archives
 - Most data in published articles: not accessible, results not replicable without the original author
 - Most data sets from NSF & NIH grants: not publicly available
- Problems even with professional archives:

Infrastructure for Quantitative Data

- Accessibility:
 - Most large data sets: in public archives
 - Most data in published articles: not accessible, results not replicable without the original author
 - Most data sets from NSF & NIH grants: not publicly available
- Problems even with professional archives:
 - Data in different archives have different identifiers

Infrastructure for Quantitative Data

- Accessibility:
 - Most large data sets: in public archives
 - Most data in published articles: not accessible, results not replicable without the original author
 - Most data sets from NSF & NIH grants: not publicly available
- Problems even with professional archives:
 - Data in different archives have different identifiers
 - One major archive renumbered all its acquisitions

Infrastructure for Quantitative Data

- Accessibility:
 - Most large data sets: in public archives
 - Most data in published articles: not accessible, results not replicable without the original author
 - Most data sets from NSF & NIH grants: not publicly available
- Problems even with professional archives:
 - Data in different archives have different identifiers
 - One major archive renumbered all its acquisitions
 - Changes to data are made; identifiers are reused or deaccessioned; old data are lost

Infrastructure for Quantitative Data

- Accessibility:
 - Most large data sets: in public archives
 - Most data in published articles: not accessible, results not replicable without the original author
 - Most data sets from NSF & NIH grants: not publicly available
- Problems even with professional archives:
 - Data in different archives have different identifiers
 - One major archive renumbered all its acquisitions
 - Changes to data are made; identifiers are reused or deaccessioned; old data are lost
- Data sets are not like books

Infrastructure for Quantitative Data

- Accessibility:
 - Most large data sets: in public archives
 - Most data in published articles: not accessible, results not replicable without the original author
 - Most data sets from NSF & NIH grants: not publicly available
- Problems even with professional archives:
 - Data in different archives have different identifiers
 - One major archive renumbered all its acquisitions
 - Changes to data are made; identifiers are reused or deaccessioned; old data are lost
- Data sets are not like books
 - Static data files (even if on the web): unreadable after a few years

Infrastructure for Quantitative Data

- Accessibility:
 - Most large data sets: in public archives
 - Most data in published articles: not accessible, results not replicable without the original author
 - Most data sets from NSF & NIH grants: not publicly available
- Problems even with professional archives:
 - Data in different archives have different identifiers
 - One major archive renumbered all its acquisitions
 - Changes to data are made; identifiers are reused or deaccessioned; old data are lost
- Data sets are not like books
 - Static data files (even if on the web): unreadable after a few years
 - When storage methods change: some data sets are lost; others have altered content!

What About a Centralized Data Access Solution?

What About a Centralized Data Access Solution?

- Highly desirable when feasible

What About a Centralized Data Access Solution?

- Highly desirable when feasible
- Works great in genomics, astronomy, etc., when data formats are universal, goals are common, and agreements are in place

What About a Centralized Data Access Solution?

- Highly desirable when feasible
- Works great in genomics, astronomy, etc., when data formats are universal, goals are common, and agreements are in place
- **Impossible** when data are heterogeneous in format, origin, size, effort needed to collect or analyze, IRB access rules, etc.

What About a Centralized Data Access Solution?

- Highly desirable when feasible
- Works great in genomics, astronomy, etc., when data formats are universal, goals are common, and agreements are in place
- **Impossible** when data are heterogeneous in format, origin, size, effort needed to collect or analyze, IRB access rules, etc.
- Why don't researchers put data in public archives?

What About a Centralized Data Access Solution?

- Highly desirable when feasible
- Works great in genomics, astronomy, etc., when data formats are universal, goals are common, and agreements are in place
- **Impossible** when data are heterogeneous in format, origin, size, effort needed to collect or analyze, IRB access rules, etc.
- Why don't researchers put data in public archives?
 - The Archive gets the credit

What About a Centralized Data Access Solution?

- Highly desirable when feasible
- Works great in genomics, astronomy, etc., when data formats are universal, goals are common, and agreements are in place
- **Impossible** when data are heterogeneous in format, origin, size, effort needed to collect or analyze, IRB access rules, etc.
- Why don't researchers put data in public archives?
 - The Archive gets the credit
 - Upon questioning: they want credit, control, and visibility

What About a Centralized Data Access Solution?

- Highly desirable when feasible
- Works great in genomics, astronomy, etc., when data formats are universal, goals are common, and agreements are in place
- **Impossible** when data are heterogeneous in format, origin, size, effort needed to collect or analyze, IRB access rules, etc.
- Why don't researchers put data in public archives?
 - The Archive gets the credit
 - Upon questioning: they want credit, control, and visibility
 - (So why don't they worry about print publishers getting all the credit?)

What About a Centralized Data Access Solution?

- Highly desirable when feasible
- Works great in genomics, astronomy, etc., when data formats are universal, goals are common, and agreements are in place
- **Impossible** when data are heterogeneous in format, origin, size, effort needed to collect or analyze, IRB access rules, etc.
- Why don't researchers put data in public archives?
 - The Archive gets the credit
 - Upon questioning: they want credit, control, and visibility
 - (So why don't they worry about print publishers getting all the credit? Lack of data citations!)

What About a Centralized Data Access Solution?

- Highly desirable when feasible
- Works great in genomics, astronomy, etc., when data formats are universal, goals are common, and agreements are in place
- **Impossible** when data are heterogeneous in format, origin, size, effort needed to collect or analyze, IRB access rules, etc.
- Why don't researchers put data in public archives?
 - The Archive gets the credit
 - Upon questioning: they want credit, control, and visibility
 - (So why don't they worry about print publishers getting all the credit? Lack of data citations!)
- We propose: technological solutions to these political and social problems

Requirements for Effective Data Sharing Infrastructure

Requirements for Effective Data Sharing Infrastructure

- **Recognition**, for authors, journals, etc. in (1) citations to data, (2) citations to associated articles, and (3) visibility on the web.

Requirements for Effective Data Sharing Infrastructure

- **Recognition**, for authors, journals, etc. in (1) citations to data, (2) citations to associated articles, and (3) visibility on the web.
- **Public Distribution**, without permission from the author

Requirements for Effective Data Sharing Infrastructure

- **Recognition**, for authors, journals, etc. in (1) citations to data, (2) citations to associated articles, and (3) visibility on the web.
- **Public Distribution**, without permission from the author
- **Authorization**: fulfill requirements the author originally met

Requirements for Effective Data Sharing Infrastructure

- **Recognition**, for authors, journals, etc. in (1) citations to data, (2) citations to associated articles, and (3) visibility on the web.
- **Public Distribution**, without permission from the author
- **Authorization**: fulfill requirements the author originally met
- **Validation**: check that data exists, without authorization

Requirements for Effective Data Sharing Infrastructure

- **Recognition**, for authors, journals, etc. in (1) citations to data, (2) citations to associated articles, and (3) visibility on the web.
- **Public Distribution**, without permission from the author
- **Authorization**: fulfill requirements the author originally met
- **Validation**: check that data exists, without authorization
- **Persistence** Decades from now. . . .

Requirements for Effective Data Sharing Infrastructure

- **Recognition**, for authors, journals, etc. in (1) citations to data, (2) citations to associated articles, and (3) visibility on the web.
- **Public Distribution**, without permission from the author
- **Authorization**: fulfill requirements the author originally met
- **Validation**: check that data exists, without authorization
- **Persistence** Decades from now. . . .
- **Verification**: data remains unchanged, even if converted

Requirements for Effective Data Sharing Infrastructure

- **Recognition**, for authors, journals, etc. in (1) citations to data, (2) citations to associated articles, and (3) visibility on the web.
- **Public Distribution**, without permission from the author
- **Authorization**: fulfill requirements the author originally met
- **Validation**: check that data exists, without authorization
- **Persistence** Decades from now. . . .
- **Verification**: data remains unchanged, even if converted from SPSS to Stata to R,

Requirements for Effective Data Sharing Infrastructure

- **Recognition**, for authors, journals, etc. in (1) citations to data, (2) citations to associated articles, and (3) visibility on the web.
- **Public Distribution**, without permission from the author
- **Authorization**: fulfill requirements the author originally met
- **Validation**: check that data exists, without authorization
- **Persistence** Decades from now. . . .
- **Verification**: data remains unchanged, even if converted from SPSS to Stata to R, from a PC to a Mac to Linux,

Requirements for Effective Data Sharing Infrastructure

- **Recognition**, for authors, journals, etc. in (1) citations to data, (2) citations to associated articles, and (3) visibility on the web.
- **Public Distribution**, without permission from the author
- **Authorization**: fulfill requirements the author originally met
- **Validation**: check that data exists, without authorization
- **Persistence** Decades from now. . . .
- **Verification**: data remains unchanged, even if converted from SPSS to Stata to R, from a PC to a Mac to Linux, and from 8 inch magnetic tape to 5.25 inch floppies to a DVD.

Requirements for Effective Data Sharing Infrastructure

- **Recognition**, for authors, journals, etc. in (1) citations to data, (2) citations to associated articles, and (3) visibility on the web.
- **Public Distribution**, without permission from the author
- **Authorization**: fulfill requirements the author originally met
- **Validation**: check that data exists, without authorization
- **Persistence** Decades from now. . . .
- **Verification**: data remains unchanged, even if converted from SPSS to Stata to R, from a PC to a Mac to Linux, and from 8 inch magnetic tape to 5.25 inch floppies to a DVD.
- **Ease of Use** Neither editors nor authors employ professional archivists

Requirements for Effective Data Sharing Infrastructure

- **Recognition**, for authors, journals, etc. in (1) citations to data, (2) citations to associated articles, and (3) visibility on the web.
- **Public Distribution**, without permission from the author
- **Authorization**: fulfill requirements the author originally met
- **Validation**: check that data exists, without authorization
- **Persistence** Decades from now. . . .
- **Verification**: data remains unchanged, even if converted from SPSS to Stata to R, from a PC to a Mac to Linux, and from 8 inch magnetic tape to 5.25 inch floppies to a DVD.
- **Ease of Use** Neither editors nor authors employ professional archivists
- **Legal Protection**:

Requirements for Effective Data Sharing Infrastructure

- **Recognition**, for authors, journals, etc. in (1) citations to data, (2) citations to associated articles, and (3) visibility on the web.
- **Public Distribution**, without permission from the author
- **Authorization**: fulfill requirements the author originally met
- **Validation**: check that data exists, without authorization
- **Persistence** Decades from now. . . .
- **Verification**: data remains unchanged, even if converted from SPSS to Stata to R, from a PC to a Mac to Linux, and from 8 inch magnetic tape to 5.25 inch floppies to a DVD.
- **Ease of Use** Neither editors nor authors employ professional archivists
- **Legal Protection**:
 - Journals have liability protection for print; none for data

Requirements for Effective Data Sharing Infrastructure

- **Recognition**, for authors, journals, etc. in (1) citations to data, (2) citations to associated articles, and (3) visibility on the web.
- **Public Distribution**, without permission from the author
- **Authorization**: fulfill requirements the author originally met
- **Validation**: check that data exists, without authorization
- **Persistence** Decades from now. . . .
- **Verification**: data remains unchanged, even if converted from SPSS to Stata to R, from a PC to a Mac to Linux, and from 8 inch magnetic tape to 5.25 inch floppies to a DVD.
- **Ease of Use** Neither editors nor authors employ professional archivists
- **Legal Protection**:
 - Journals have liability protection for print; none for data
 - *If you put data on the web without IRB approval, you are violating federal regulations*

Requirements for Effective Data Sharing Infrastructure

- **Recognition**, for authors, journals, etc. in (1) citations to data, (2) citations to associated articles, and (3) visibility on the web.
- **Public Distribution**, without permission from the author
- **Authorization**: fulfill requirements the author originally met
- **Validation**: check that data exists, without authorization
- **Persistence** Decades from now. . . .
- **Verification**: data remains unchanged, even if converted from SPSS to Stata to R, from a PC to a Mac to Linux, and from 8 inch magnetic tape to 5.25 inch floppies to a DVD.
- **Ease of Use** Neither editors nor authors employ professional archivists
- **Legal Protection**:
 - Journals have liability protection for print; none for data
 - *If you put data on the web without IRB approval, you are violating federal regulations*
 - (IRB approval must be for data distribution, not merely for the study)

Requirements for Effective Data Sharing Infrastructure

- **Recognition**, for authors, journals, etc. in (1) citations to data, (2) citations to associated articles, and (3) visibility on the web.
- **Public Distribution**, without permission from the author
- **Authorization**: fulfill requirements the author originally met
- **Validation**: check that data exists, without authorization
- **Persistence** Decades from now. . . .
- **Verification**: data remains unchanged, even if converted from SPSS to Stata to R, from a PC to a Mac to Linux, and from 8 inch magnetic tape to 5.25 inch floppies to a DVD.
- **Ease of Use** Neither editors nor authors employ professional archivists
- **Legal Protection**:
 - Journals have liability protection for print; none for data
 - *If you put data on the web without IRB approval, you are violating federal regulations*
 - (IRB approval must be for data distribution, not merely for the study)
 - Solution must not require lawyers (we've automated the IRB)

Rules for Citing Printed Matter

Rules for Citing Printed Matter

Kim, Jae-On, Norman Nie, and Sidney Verba. 1977. "A Note on Factor Analyzing Dichotomous Variables: The Case of Political Participation," Political Methodology, Vol. 4: No. 2 (Spring): Pp. 39–62.

Rules for Citing Printed Matter

Kim, Jae-On, Norman Nie, and Sidney Verba. 1977. "A Note on Factor Analyzing Dichotomous Variables: The Case of Political Participation," Political Methodology, Vol. 4: No. 2 (Spring): Pp. 39–62.

First author (last name first)

Rules for Citing Printed Matter

Kim, Jae-On, *Norman Nie*, and Sidney Verba. 1977. "A Note on Factor Analyzing Dichotomous Variables: The Case of Political Participation," *Political Methodology*, Vol. 4: No. 2 (Spring): Pp. 39–62.

Second author

Rules for Citing Printed Matter

Kim, Jae-On, Norman Nie, and *Sidney Verba*. 1977. "A Note on Factor Analyzing Dichotomous Variables: The Case of Political Participation," *Political Methodology*, Vol. 4: No. 2 (Spring): Pp. 39–62.

Third author

Rules for Citing Printed Matter

Kim, Jae-On, Norman Nie, and Sidney Verba. 1977. "A Note on Factor Analyzing Dichotomous Variables: The Case of Political Participation," *Political Methodology*, Vol. 4: No. 2 (Spring): Pp. 39–62.

Year

Rules for Citing Printed Matter

Kim, Jae-On, Norman Nie, and Sidney Verba. 1977. "*A Note on Factor Analyzing Dichotomous Variables: The Case of Political Participation*," *Political Methodology*, Vol. 4: No. 2 (Spring): Pp. 39–62.

Article title

Rules for Citing Printed Matter

Kim, Jae-On, Norman Nie, and Sidney Verba. 1977. "A Note on Factor Analyzing Dichotomous Variables: The Case of Political Participation," *Political Methodology*, Vol. 4: No. 2 (Spring): Pp. 39–62.

Journal (no longer exists)

Rules for Citing Printed Matter

Kim, Jae-On, Norman Nie, and Sidney Verba. 1977. "A Note on Factor Analyzing Dichotomous Variables: The Case of Political Participation," Political Methodology, Vol. 4: No. 2 (Spring): Pp. 39–62.

Volume number

Rules for Citing Printed Matter

Kim, Jae-On, Norman Nie, and Sidney Verba. 1977. "A Note on Factor Analyzing Dichotomous Variables: The Case of Political Participation," Political Methodology, Vol. 4: No. 2 (Spring): Pp. 39–62.

Issue number

Rules for Citing Printed Matter

Kim, Jae-On, Norman Nie, and Sidney Verba. 1977. "A Note on Factor Analyzing Dichotomous Variables: The Case of Political Participation," Political Methodology, Vol. 4: No. 2 (Spring): Pp. 39–62.

Season

Rules for Citing Printed Matter

Kim, Jae-On, Norman Nie, and Sidney Verba. 1977. "A Note on Factor Analyzing Dichotomous Variables: The Case of Political Participation," Political Methodology, Vol. 4: No. 2 (Spring): Pp. 39–62.

Pages

Rules for Citing Printed Matter

Kim, Jae-On, Norman Nie, and Sidney Verba. 1977. "A Note on Factor Analyzing Dichotomous Variables: The Case of Political Participation," Political Methodology, Vol. 4: No. 2 (Spring): Pp. 39–62.

Special formatting codes

Rules for Citing Printed Matter

Kim, Jae-On, Norman Nie, and Sidney Verba. 1977. "A Note on Factor Analyzing Dichotomous Variables: The Case of Political Participation," Political Methodology, Vol. 4: No. 2 (Spring): Pp. 39–62.

Special indentation

Rules for Citing Printed Matter

Kim, Jae-On, Norman Nie, and Sidney Verba. 1977. "A Note on Factor Analyzing Dichotomous Variables: The Case of Political Participation," Political Methodology, Vol. 4: No. 2 (Spring): Pp. 39–62.

Citations: rule-based, precise, redundant

Rules for Citing Printed Matter

Kim, Jae-On, Norman Nie, and Sidney Verba. 1977. "A Note on Factor Analyzing Dichotomous Variables: The Case of Political Participation," Political Methodology, Vol. 4: No. 2 (Spring): Pp. 39–62.

Print Citations Work: authors don't think publishers get all the credit; cited articles can be found; copyeditors don't need to see the original to know it exists; the link from citation to print persists

A New Citation Standard for Numeric Data

Sidney Verba, 1998, "Political Participation Data", [hdl:1902.4/00754](https://hdl.handle.net/1902.4/00754),
UNF:3:6:ZNQRI14053UZq389x0Bffg?==

A New Citation Standard for Numeric Data

Sidney Verba, 1998, "Political Participation Data", [hdl:1902.4/00754](https://hdl.handle.net/1902.4/00754),
UNF:3:6:ZNQRI14053UZq389x0Bffg?==

 Author

A New Citation Standard for Numeric Data

Sidney Verba, 1998, "Political Participation Data", [hdl:1902.4/00754](https://hdl.handle.net/1902.4/00754),
UNF:3:6:ZNQRI14053UZq389x0Bffg?==

① Author

② Year

A New Citation Standard for Numeric Data

Sidney Verba, 1998, “**Political Participation Data**”, [hdl:1902.4/00754](https://hdl.handle.net/1902.4/00754),
UNF:3:6:ZNQRI14053UZq389x0Bffg?==

- 1 Author
- 2 Year
- 3 **Title**

A New Citation Standard for Numeric Data

Sidney Verba, 1998, "Political Participation Data", [hdl:1902.4/00754](https://hdl.handle.net/1902.4/00754),
UNF:3:6:ZNQRI14053UZq389x0Bffg?==

- 1 Author
- 2 Year
- 3 Title
- 4 Unique Global Identifier: will work after URLs stop working

A New Citation Standard for Numeric Data

Sidney Verba, 1998, "Political Participation Data", [hdl:1902.4/00754](https://hdl.handle.net/1902.4/00754),
UNF:3:6:ZNQRI14053UZq389x0Bffg?==

- 1 Author
- 2 Year
- 3 Title
- 4 Unique Global Identifier: will work after URLs stop working
- 5 Linked to a Bridge Service (presently a URL:
<http://id.thedata.org/hdl%3A1902.4%2F00754>)

A New Citation Standard for Numeric Data

Sidney Verba, 1998, "Political Participation Data", [hdl:1902.4/00754](https://hdl.handle.net/1902.4/00754),
UNF:3:6:ZNQRI14053UZq389x0Bffg?==

- 1 Author
- 2 Year
- 3 Title
- 4 Unique Global Identifier: will work after URLs stop working
- 5 Linked to a Bridge Service (presently a URL:
<http://id.thedata.org/hdl%3A1902.4%2F00754>)
- 6 **Universal Numeric Fingerprint (UNF)**

A New Citation Standard for Numeric Data

Sidney Verba, 1998, "Political Participation Data", [hdl:1902.4/00754](https://hdl.handle.net/1902.4/00754),
UNF:3:6:ZNQRI14053UZq389x0Bffg?== **Annals of Applied Statistics**
[Distributor];

- 1 Author
- 2 Year
- 3 Title
- 4 Unique Global Identifier: will work after URLs stop working
- 5 Linked to a Bridge Service (presently a URL:
<http://id.thedata.org/hdl%3A1902.4%2F00754>)
- 6 Universal Numeric Fingerprint (UNF)
- 7 **Standard rules for adding citation elements**

A New Citation Standard for Numeric Data

Sidney Verba, 1998, "Political Participation Data", [hdl:1902.4/00754](https://hdl.handle.net/1902.4/00754),
UNF:3:6:ZNQRI14053UZq389x0Bffg?== **Annals of Applied Statistics**
[Distributor]; NORC [Producer].

- 1 Author
- 2 Year
- 3 Title
- 4 Unique Global Identifier: will work after URLs stop working
- 5 Linked to a Bridge Service (presently a URL:
<http://id.thedata.org/hdl%3A1902.4%2F00754>)
- 6 Universal Numeric Fingerprint (UNF)
- 7 **Standard rules for adding citation elements**

Data to Universal Numeric Fingerprints

Data to Universal Numeric Fingerprints

$$\begin{pmatrix} 1 & 4 & 4 & 21 & \dots & 121 \\ 1 & 2 & 2 & 91 & \dots & 212 \\ 1 & 9 & 2 & 72 & \dots & 104 \\ 0 & 2 & 2 & 2 & \dots & 321 \\ 1 & 6 & 2 & 12 & \dots & 204 \\ 1 & 9 & 4 & 52 & \dots & 311 \\ 0 & 3 & 2 & 23 & \dots & 92 \\ 0 & 2 & 5 & 91 & \dots & 212 \\ 0 & 5 & 8 & 91 & \dots & 91 \\ 1 & 9 & 1 & 72 & \dots & 104 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 2 & 2 & 91 & \dots & 212 \end{pmatrix}$$

Data to Universal Numeric Fingerprints

$$\begin{pmatrix} 1 & 4 & 4 & 21 & \dots & 121 \\ 1 & 2 & 2 & 91 & \dots & 212 \\ 1 & 9 & 2 & 72 & \dots & 104 \\ 0 & 2 & 2 & 2 & \dots & 321 \\ 1 & 6 & 2 & 12 & \dots & 204 \\ 1 & 9 & 4 & 52 & \dots & 311 \\ 0 & 3 & 2 & 23 & \dots & 92 \\ 0 & 2 & 5 & 91 & \dots & 212 \\ 0 & 5 & 8 & 91 & \dots & 91 \\ 1 & 9 & 1 & 72 & \dots & 104 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 2 & 2 & 91 & \dots & 212 \end{pmatrix}$$

\Rightarrow ZNQRI14053UZq389x0Bffg?==

Advantages of UNFs

Advantages of UNFs

- UNF is **calculated from the content** not the file:

.

Advantages of UNFs

- UNF is **calculated from the content** not the file: Its the **Same UNF** regardless of changes in computer hardware,

Advantages of UNFs

- UNF is **calculated from the content** not the file: Its the **Same UNF** regardless of changes in computer hardware, storage medium,

.

Advantages of UNFs

- UNF is **calculated from the content** not the file: Its the **Same UNF** regardless of changes in computer hardware, storage medium, operating system,

Advantages of UNFs

- UNF is **calculated from the content** not the file: Its the **Same UNF** regardless of changes in computer hardware, storage medium, operating system, statistical software,

Advantages of UNFs

- UNF is **calculated from the content** not the file: Its the **Same UNF** regardless of changes in computer hardware, storage medium, operating system, statistical software, database,

Advantages of UNFs

- UNF is **calculated from the content** not the file: Its the **Same UNF** regardless of changes in computer hardware, storage medium, operating system, statistical software, database, or spreadsheet software.

Advantages of UNFs

- UNF is **calculated from the content** not the file: Its the **Same UNF** regardless of changes in computer hardware, storage medium, operating system, statistical software, database, or spreadsheet software.
- Cryptographic technology: **any** change in data content changes the UNF. (cannot tinker after the fact!)

Advantages of UNFs

- UNF is **calculated from the content** not the file: Its the **Same UNF** regardless of changes in computer hardware, storage medium, operating system, statistical software, database, or spreadsheet software.
- Cryptographic technology: **any** change in data content changes the UNF. (cannot tinker after the fact!)
- Noninvertible properties

Advantages of UNFs

- UNF is **calculated from the content** not the file: Its the **Same UNF** regardless of changes in computer hardware, storage medium, operating system, statistical software, database, or spreadsheet software.
- Cryptographic technology: **any** change in data content changes the UNF. (cannot tinker after the fact!)
- Noninvertible properties
 - UNFs convey no information about data content

Advantages of UNFs

- UNF is **calculated from the content** not the file: Its the **Same UNF** regardless of changes in computer hardware, storage medium, operating system, statistical software, database, or spreadsheet software.
- Cryptographic technology: **any** change in data content changes the UNF. (cannot tinker after the fact!)
- Noninvertible properties
 - UNFs convey no information about data content
 - OK to distribute for highly sensitive, confidential, or proprietary data

Advantages of UNFs

- UNF is **calculated from the content** not the file: Its the **Same UNF** regardless of changes in computer hardware, storage medium, operating system, statistical software, database, or spreadsheet software.
- Cryptographic technology: **any** change in data content changes the UNF. (cannot tinker after the fact!)
- Noninvertible properties
 - UNFs convey no information about data content
 - OK to distribute for highly sensitive, confidential, or proprietary data
 - Copyeditor can validate data's existence even without authorization

Advantages of UNFs

- UNF is **calculated from the content** not the file: Its the **Same UNF** regardless of changes in computer hardware, storage medium, operating system, statistical software, database, or spreadsheet software.
- Cryptographic technology: **any** change in data content changes the UNF. (cannot tinker after the fact!)
- Noninvertible properties
 - UNFs convey no information about data content
 - OK to distribute for highly sensitive, confidential, or proprietary data
 - Copyeditor can validate data's existence even without authorization
- The citation refers to **one specific data set** that can't **ever** be altered, even if journal doesn't keep a copy

Advantages of UNFs

- UNF is **calculated from the content** not the file: Its the **Same UNF** regardless of changes in computer hardware, storage medium, operating system, statistical software, database, or spreadsheet software.
- Cryptographic technology: **any** change in data content changes the UNF. (cannot tinker after the fact!)
- Noninvertible properties
 - UNFs convey no information about data content
 - OK to distribute for highly sensitive, confidential, or proprietary data
 - Copyeditor can validate data's existence even without authorization
- The citation refers to **one specific data set** that can't **ever** be altered, even if journal doesn't keep a copy
- Future researchers can quickly check that they have the same data as used by the author: merely recalculate the UNF

Web 2.0 Terminology

Web 2.0 Terminology

- **Software:** find CD, install locally,

Web 2.0 Terminology

- **Software:** find CD, install locally, hit next,

Web 2.0 Terminology

- **Software:** find CD, install locally, hit next, hit next,

- **Software:** find CD, install locally, hit next, hit next, hit next. . .

Web 2.0 Terminology

- **Software**: find CD, install locally, hit next, hit next, hit next. . .
- **Web application software**: no installation; load web browser and run (Dataverse Network Software)

Web 2.0 Terminology

- **Software**: find CD, install locally, hit next, hit next, hit next. . .
- **Web application software**: no installation; load web browser and run (Dataverse Network Software)
- **Host**: The computers where the web application software runs (universities, archives, libraries)

Web 2.0 Terminology

- **Software**: find CD, install locally, hit next, hit next, hit next. . .
- **Web application software**: no installation; load web browser and run (Dataverse Network Software)
- **Host**: The computers where the web application software runs (universities, archives, libraries)
- **Virtual host**: Where the web application software *seems* to run, but does not (web sites of: authors, journals, granting agencies, research centers, universities, scholarly organizations, etc.)

Annals of Applied Applications
The Journal for those who truly apply themselves

U.S. open-source group opens chapter in Europe
by Peter Judge
January 30, 2008, 6:37 AM PST

"More open-source projects have originated in Europe than anywhere else in the world," said Bertrand Dard, chief executive officer of data-integration specialist Tarevi, a founding member of the Open Solutions Alliance. Founded a year ago, the OSA has had a U.S. focus until now.

The European chapter of the OSA will be formally incorporated in the next 90 days, and will then look for interoperability work required by European users.

"When the OSA got started, we saw our mission as a global one, but the critical mass of activity so far has been in the U.S.," said Dominic Sartorio, OSA president and senior director of product management at services company SpliceSource. "We were approached two months ago by a group of companies in Europe thinking of forming a group like ours but focused on Europe. This led us to think that while we are trying to achieve results globally, there are some parts of the world where we are doing less well."

Sartorio sees 2008 as turning point for open source. In late 2007, he promised that OSA would "out-Microsoft Microsoft" in response to a user survey that revealed a demand for interoperability among open-source tools.

"We have encountered other groups with a similar mission, like the U.K.'s OpenForum, but they tend to focus on pure advocacy," said Sartorio. "The OSA focuses on business applications and their interoperability."

Open-source companies need interoperability because they are usually small and focused, he said. When they get bought by larger companies, this helps them integrate, but not all open-source companies have that option.

Open source is growing rapidly in Europe. In France, total spending on open-source services and products jumped 66 percent to \$1.07 billion in 2007, according to Paris-based analysts Pierre Audoin Consultants.

Other OSA chapters are expected elsewhere in the world. "Our chapter system is intended to scale around the world," said Sartorio. "I would expect by this time next year we will have chapters up and running in other regions enjoying strong open-source adoption, including Asia and Latin America" ours but focused on Europe. This led us to think that while we are trying to achieve results globally, there are some parts of the world where we are doing less well."

Sartorio sees 2008 as turning point for open source. In late 2007, he promised that OSA would "out-Microsoft Microsoft" in response to a user survey that revealed a demand for interoperability among open-source tools.

About AAA
Editorial Board
Contact Info
Current Issue
Letters
Dataverse

Your web site

Annals of Applied Applications
The Journal for those who truly apply themselves

All IQSS Databases > Annals of Applied Applications

Search/Browse User Guide Site Map Contact Us Login Harvard Athlete

STUDIES

Search Cataloging Information for Search Advanced Search

- Annals of Applied Applications
 - Supplements: Volume Issues
 - Supplements: Volume 1, Number 1(2007), pp. 1-283
 - Supplements: Volume Issues
 - Supplements: Volume 1, Number 2(2007), pp. 284-506
 - Supplements: Volume Issues
 - Supplements: Volume 1, Number 3(2007), pp. 507-785
 - Supplements: Volume Issues
 - Supplements: Volume 1, Number 4(2007), pp. 785-983
 - Volume 2, 2008
 - Issue 1
 - Issue 2
 - Issue 3

About AAA
Editorial Board
Contact Info
Current Issue
Letters
Dataverse

Your dataverse branded as your web site but served by the Dataverse Network, therefore re-quiring no local installation and providing an enormous array of services

Your DWN Google

<http://dwn.is.harvard.edu/dwn/dvbyes>



Annals of Applied Applications

The journal for those who truly apply themselves

All IQSS Databases >
Annals of Applied Applications

Search/Browse User Guide Site Map Contact Us Log in Harvard Affiliates

REPLICATION DATA FOR: IMPROVING FORECASTS OF STATE FAILURE

Cataloging Information Documentation, Data and Analysis

Citation Information

Gary King, Langhe Zeng, 2001, "Replication data for Improving Forecasts of State Failure", doi:10.1.1.RPQQDQNR.UHF3.CBafGhbaBafYuhZWWWtQW Mummy Research Archive (Distributor)

How to Cite

Study Global Id: hdl:1902.1/RPQQDQNR

Authors: Gary King, Langhe Zeng

Production Date: 2001

Distributor: Mummy Research Archive **MRA**

Distributor Contact: mm_suzoon@help.hinds.harvard.edu

Deposit Date: 2006

Replication For: King, Gary, Zeng, Langhe, 2001, "Improving Forecasts of State Failure", World Politics, Vol. 52, No. 4, 829-36. <http://gking.harvard.edu/repdata/priv-acc.shtml> article available here.


Provenance: Gary King Database

Abstract and Scope

We offer the first independent scholarly evaluation of the claims, forecasts, and causal inferences of the State Failure Task Force and their efforts to forecast when states will fail. State failure refers to the collapse of the authority of the central government to impose order, as in civil wars, revolutionary wars,

Your DWN Google

<http://dwn.is.harvard.edu/dwn/dvbyes>



Annals of Applied Applications

The journal for those who truly apply themselves

All IQSS Databases >
Annals of Applied Applications Available in Dataverse Network

Search/Browse User Guide Site Map Contact Us Log in Harvard Affiliates

REPLICATION DATA FOR: IMPROVING FORECASTS OF STATE FAILURE

Cataloging Information Documentation, Data and Analysis

Download all files in a single archive file (files that you cannot access will not be downloaded)

File Name	Description	Owner	Variables	Type	Control
1. Documentation					
improvingforecasts.pdf	Article related to this study: Improving Forecasts of State Failure			application/pdf	
2. Replication Data					
nr01.aspc	Sample spec file for committee nr's				
nr.dat	For use with committee nr's, ASCII data file				
nr.tab	For use with committee nr's	7190	11	Tab-delimited	
replication.txt	Additional document describing the format of the replication files				
3. Recoded Data					
names.dta	Country names, in original Stata format				
names.tab	Country Names Data	8980	3	Tab-delimited	
sp111r01.dta	Recoded Full Data in Original Format, comma delimited text file				
4. Original Data					
repdata.dta	Documentation for Original				

Year DYN

http://dyn.us.harvard.edu/dyn/dyn/

Annals of Applied Applications
The Journal for those who truly apply themselves

All IQSS Databases >
 Annals of Applied Applications **Dataverse Network**

Search Datasets User Guides Site Map Contact Us Log In Harvard Alliance

REPLICATION DATA FOR CONSTITUENCY SERVICE AND INCLINENCY ADVANTAGE [Back to Study](#)
 DATA FILE: CONSTITUENCYSERVICE.TAB

Download Subset Subset and Records Descriptive Statistics **Advanced Statistical Analysis**

Selected Variables

Choose a Statistical Model

Choose a Statistical Model

Categorical Data Analysis
 Cross-Tabulation

Ecological Inference models
 Hierarchical Multinomial-Dirichlet Ecological Inference Model for R x C Tables

Select Event Count Models

variables all

- Negative Binomial Regression for Event Count Dependent Variables
- Poisson Regression for Event Count Dependent Variables
- Generalized Additive Model for Event Count Dependent Variables
- General Estimating Equation for Poisson Regression
- Social Network Poisson Regression for Event Count Dependent Variables

Models for Continuous Bounded Dependent Variables

- Cox Proportional Hazard Regression for Duration Dependent Variables
- Exponential Regression for Duration Dependent Variables
- Gamma Regression for Continuous, Positive Dependent Variables
- General Estimating Equation for Gamma Regression

Discrete

<input type="checkbox"/>	CF	state level dummy variables	11
<input type="checkbox"/>	DE	state level dummy variables	11
<input type="checkbox"/>	IL	state level dummy variables	11
<input type="checkbox"/>	MI	state level dummy variables	11
<input type="checkbox"/>	MO	state level dummy variables	11
<input type="checkbox"/>	SAL	salary	11
<input type="checkbox"/>	CO	state level dummy variables	11

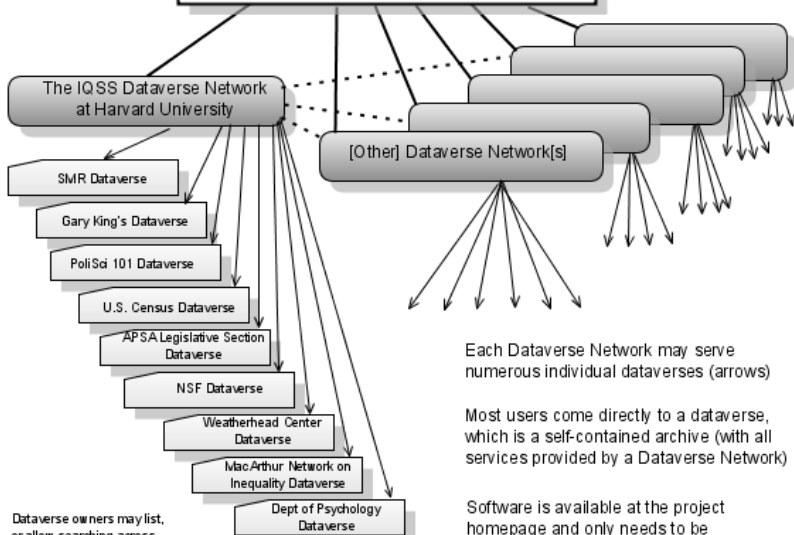
About AAA
 Editorial Board
 Contact Info
 Current Issue
 Letters
 Dataverse

Your web site

Your dataverse branded as your web site but served by the Dataverse Network, therefore re-quiring no local installation and providing an enormous array of services

The Dataverse Network Project Homepage (<http://TheData.org>)

Dataverse Networks may harvest metadata from each other (dashed lines)



Dataverse owners may list, or allow searching across, other dataverses or the data sets in them

Each Dataverse Network may serve numerous individual dataverses (arrows)

Most users come directly to a dataverse, which is a self-contained archive (with all services provided by a Dataverse Network)

Software is available at the project homepage and only needs to be installed to establish a Dataverse Network. Dataverses are virtual hosts.

A Journal Dataverse for a Replication Data Archive

A Journal Dataverse for a Replication Data Archive

- **Full service virtual archive**, with numerous data services (citation, metadata, archiving, subsetting, conversion, translation, analysis, ...)

A Journal Dataverse for a Replication Data Archive

- **Full service virtual archive**, with numerous data services (citation, metadata, archiving, subsetting, conversion, translation, analysis, ...)
- **Hierarchical list of data**: year/volume/issue/article/dataset

A Journal Dataverse for a Replication Data Archive

- **Full service virtual archive**, with numerous data services (citation, metadata, archiving, subsetting, conversion, translation, analysis, ...)
- **Hierarchical list of data**: year/volume/issue/article/dataset
- **Branded as yours**: with the look and feel of your site

A Journal Dataverse for a Replication Data Archive

- **Full service virtual archive**, with numerous data services (citation, metadata, archiving, subsetting, conversion, translation, analysis, . . .)
- **Hierarchical list of data**: year/volume/issue/article/dataset
- **Branded as yours**: with the look and feel of your site
- **Easy to setup**: give DVN your style, and include a link to your new dataverse

A Journal Dataverse for a Replication Data Archive

- **Full service virtual archive**, with numerous data services (citation, metadata, archiving, subsetting, conversion, translation, analysis, . . .)
- **Hierarchical list of data**: year/volume/issue/article/dataset
- **Branded as yours**: with the look and feel of your site
- **Easy to setup**: give DVN your style, and include a link to your new dataverse
- **Easy to manage**: no software or hardware installation, backups, worry about archiving standards, or data format transations; still exists if you move; easy to rebrand

A Journal Dataverse for a Replication Data Archive

- **Full service virtual archive**, with numerous data services (citation, metadata, archiving, subsetting, conversion, translation, analysis, ...)
- **Hierarchical list of data**: year/volume/issue/article/dataset
- **Branded as yours**: with the look and feel of your site
- **Easy to setup**: give DVN your style, and include a link to your new dataverse
- **Easy to manage**: no software or hardware installation, backups, worry about archiving standards, or data format transations; still exists if you move; easy to rebrand
- **Does not disrupt workflow**: copyeditor ensures data is cited; author is given password to upload data. Everything else is self-service

A Journal Dataverse for a Replication Data Archive

- **Full service virtual archive**, with numerous data services (citation, metadata, archiving, subsetting, conversion, translation, analysis, ...)
- **Hierarchical list of data**: year/volume/issue/article/dataset
- **Branded as yours**: with the look and feel of your site
- **Easy to setup**: give DVN your style, and include a link to your new dataverse
- **Easy to manage**: no software or hardware installation, backups, worry about archiving standards, or data format transations; still exists if you move; easy to rebrand
- **Does not disrupt workflow**: copyeditor ensures data is cited; author is given password to upload data. Everything else is self-service
- **High acceptability**: experiments indicate $> 90\%$ uptake for authors (without publication at stake)

A Journal Dataverse for a Replication Data Archive

- **Full service virtual archive**, with numerous data services (citation, metadata, archiving, subsetting, conversion, translation, analysis, ...)
- **Hierarchical list of data**: year/volume/issue/article/dataset
- **Branded as yours**: with the look and feel of your site
- **Easy to setup**: give DVN your style, and include a link to your new dataverse
- **Easy to manage**: no software or hardware installation, backups, worry about archiving standards, or data format transations; still exists if you move; easy to rebrand
- **Does not disrupt workflow**: copyeditor ensures data is cited; author is given password to upload data. Everything else is self-service
- **High acceptability**: experiments indicate $> 90\%$ uptake for authors (without publication at stake)
- **Reuse**: data may appear on author's dataverse too

A Journal Dataverse for a Replication Data Archive

- **Full service virtual archive**, with numerous data services (citation, metadata, archiving, subsetting, conversion, translation, analysis, ...)
- **Hierarchical list of data**: year/volume/issue/article/dataset
- **Branded as yours**: with the look and feel of your site
- **Easy to setup**: give DVN your style, and include a link to your new dataverse
- **Easy to manage**: no software or hardware installation, backups, worry about archiving standards, or data format transations; still exists if you move; easy to rebrand
- **Does not disrupt workflow**: copyeditor ensures data is cited; author is given password to upload data. Everything else is self-service
- **High acceptability**: experiments indicate $> 90\%$ uptake for authors (without publication at stake)
- **Reuse**: data may appear on author's dataverse too
- **Results**: Journals with replication policies have three times the impact factor! (with dataverse, it should be more)

Dataverse Uses

Dataverse Uses

- Journals, for replication data archives

Dataverse Uses

- Journals, for replication data archives
- Authors, for their own data

Dataverse Uses

- Journals, for replication data archives
- Authors, for their own data
- Future Researchers: browse or search for a dataverse or dataset; forward citation search; verification via UNFs; subsetting; read metadata, abstract, & documentation; check for new versions; translate format; statistical analyses; download

Dataverse Uses

- Journals, for replication data archives
- Authors, for their own data
- Future Researchers: browse or search for a dataverse or dataset; forward citation search; verification via UNFs; subsetting; read metadata, abstract, & documentation; check for new versions; translate format; statistical analyses; download
- Teachers, a list or for in depth analysis

Dataverse Uses

- Journals, for replication data archives
- Authors, for their own data
- Future Researchers: browse or search for a dataverse or dataset; forward citation search; verification via UNFs; subsetting; read metadata, abstract, & documentation; check for new versions; translate format; statistical analyses; download
- Teachers, a list or for in depth analysis
- Sections of scholarly organizations, to organize existing data

Dataverse Uses

- Journals, for replication data archives
- Authors, for their own data
- Future Researchers: browse or search for a dataverse or dataset; forward citation search; verification via UNFs; subsetting; read metadata, abstract, & documentation; check for new versions; translate format; statistical analyses; download
- Teachers, a list or for in depth analysis
- Sections of scholarly organizations, to organize existing data
- Granting agencies

Dataverse Uses

- Journals, for replication data archives
- Authors, for their own data
- Future Researchers: browse or search for a dataverse or dataset; forward citation search; verification via UNFs; subsetting; read metadata, abstract, & documentation; check for new versions; translate format; statistical analyses; download
- Teachers, a list or for in depth analysis
- Sections of scholarly organizations, to organize existing data
- Granting agencies
- Research centers

Dataverse Uses

- Journals, for replication data archives
- Authors, for their own data
- Future Researchers: browse or search for a dataverse or dataset; forward citation search; verification via UNFs; subsetting; read metadata, abstract, & documentation; check for new versions; translate format; statistical analyses; download
- Teachers, a list or for in depth analysis
- Sections of scholarly organizations, to organize existing data
- Granting agencies
- Research centers
- Major Research Projects

Dataverse Uses

- Journals, for replication data archives
- Authors, for their own data
- Future Researchers: browse or search for a dataverse or dataset; forward citation search; verification via UNFs; subsetting; read metadata, abstract, & documentation; check for new versions; translate format; statistical analyses; download
- Teachers, a list or for in depth analysis
- Sections of scholarly organizations, to organize existing data
- Granting agencies
- Research centers
- Major Research Projects
- Academic departments, universities, data centers, libraries

Dataverse Uses

- Journals, for replication data archives
- Authors, for their own data
- Future Researchers: browse or search for a dataverse or dataset; forward citation search; verification via UNFs; subsetting; read metadata, abstract, & documentation; check for new versions; translate format; statistical analyses; download
- Teachers, a list or for in depth analysis
- Sections of scholarly organizations, to organize existing data
- Granting agencies
- Research centers
- Major Research Projects
- Academic departments, universities, data centers, libraries
- Data archives

Dataverse Uses

- Journals, for replication data archives
- Authors, for their own data
- Future Researchers: browse or search for a dataverse or dataset; forward citation search; verification via UNFs; subsetting; read metadata, abstract, & documentation; check for new versions; translate format; statistical analyses; download
- Teachers, a list or for in depth analysis
- Sections of scholarly organizations, to organize existing data
- Granting agencies
- Research centers
- Major Research Projects
- Academic departments, universities, data centers, libraries
- Data archives
- Using DVN tools with outside data

The Universe of Data meets the Universe of Methods

The Universe of Data meets the Universe of Methods

- R Project for Statistical Computing

The Universe of Data meets the Universe of Methods

- R Project for Statistical Computing
 - nearly 1000 packages; most new methods appear in R first

The Universe of Data meets the Universe of Methods

- R Project for Statistical Computing
 - nearly 1000 packages; most new methods appear in R first
 - Highly diverse examples, syntax, documentation, and quality

- **R Project for Statistical Computing**
 - nearly 1000 packages; most new methods appear in R first
 - Highly diverse examples, syntax, documentation, and quality
 - Difficult for statistics-types; harder for applied researchers

The Universe of Data meets the Universe of Methods

- **R Project for Statistical Computing**
 - nearly 1000 packages; most new methods appear in R first
 - Highly diverse examples, syntax, documentation, and quality
 - Difficult for statistics-types; harder for applied researchers
- **Zelig: Everyone's Statistical Software**

The Universe of Data meets the Universe of Methods

- **R Project for Statistical Computing**
 - nearly 1000 packages; most new methods appear in R first
 - Highly diverse examples, syntax, documentation, and quality
 - Difficult for statistics-types; harder for applied researchers
- **Zelig: Everyone's Statistical Software**
 - An ontology we developed of almost all statistical methods

The Universe of Data meets the Universe of Methods

- **R Project for Statistical Computing**
 - nearly 1000 packages; most new methods appear in R first
 - Highly diverse examples, syntax, documentation, and quality
 - Difficult for statistics-types; harder for applied researchers
- **Zelig: Everyone's Statistical Software**
 - An ontology we developed of almost all statistical methods
 - Users incorporate original packages via simple bridge functions via a simple model description language

The Universe of Data meets the Universe of Methods

- **R Project for Statistical Computing**

- nearly 1000 packages; most new methods appear in R first
- Highly diverse examples, syntax, documentation, and quality
- Difficult for statistics-types; harder for applied researchers

- **Zelig: Everyone's Statistical Software**

- An ontology we developed of almost all statistical methods
- Users incorporate original packages via simple bridge functions via a simple model description language
- Result: Unified Syntax, the same 3 commands to use any method

- **R Project for Statistical Computing**

- nearly 1000 packages; most new methods appear in R first
- Highly diverse examples, syntax, documentation, and quality
- Difficult for statistics-types; harder for applied researchers

- **Zelig: Everyone's Statistical Software**

- An ontology we developed of almost all statistical methods
- Users incorporate original packages via simple bridge functions via a simple model description language
- Result: Unified Syntax, the same 3 commands to use any method
- Automatically generated Graphical User Interface with all the world's methods

The Universe of Data meets the Universe of Methods

- **R Project for Statistical Computing**

- nearly 1000 packages; most new methods appear in R first
- Highly diverse examples, syntax, documentation, and quality
- Difficult for statistics-types; harder for applied researchers

- **Zelig: Everyone's Statistical Software**

- An ontology we developed of almost all statistical methods
- Users incorporate original packages via simple bridge functions via a simple model description language
- Result: Unified Syntax, the same 3 commands to use any method
- Automatically generated Graphical User Interface with all the world's methods

- **R + Zelig + Dataverse Network**

The Universe of Data meets the Universe of Methods

- **R Project for Statistical Computing**

- nearly 1000 packages; most new methods appear in R first
- Highly diverse examples, syntax, documentation, and quality
- Difficult for statistics-types; harder for applied researchers

- **Zelig: Everyone's Statistical Software**

- An ontology we developed of almost all statistical methods
- Users incorporate original packages via simple bridge functions via a simple model description language
- Result: Unified Syntax, the same 3 commands to use any method
- Automatically generated Graphical User Interface with all the world's methods

- **R + Zelig + Dataverse Network**

- Greatly reduced time from methods development to use

The Universe of Data meets the Universe of Methods

- **R Project for Statistical Computing**
 - nearly 1000 packages; most new methods appear in R first
 - Highly diverse examples, syntax, documentation, and quality
 - Difficult for statistics-types; harder for applied researchers
- **Zelig: Everyone's Statistical Software**
 - An ontology we developed of almost all statistical methods
 - Users incorporate original packages via simple bridge functions via a simple model description language
 - Result: Unified Syntax, the same 3 commands to use any method
 - Automatically generated Graphical User Interface with all the world's methods
- **R + Zelig + Dataverse Network**
 - Greatly reduced time from methods development to use
 - Easy for applied researchers even if non-programmers

The Universe of Data meets the Universe of Methods

- **R Project for Statistical Computing**

- nearly 1000 packages; most new methods appear in R first
- Highly diverse examples, syntax, documentation, and quality
- Difficult for statistics-types; harder for applied researchers

- **Zelig: Everyone's Statistical Software**

- An ontology we developed of almost all statistical methods
- Users incorporate original packages via simple bridge functions via a simple model description language
- Result: Unified Syntax, the same 3 commands to use any method
- Automatically generated Graphical User Interface with all the world's methods

- **R + Zelig + Dataverse Network**

- Greatly reduced time from methods development to use
- Easy for applied researchers even if non-programmers
- Can save R script for replication or further analysis

- Web application software

- **Web application software**
 - Pros: easy to use, no installation costs

- **Web application software**
 - Pros: easy to use, no installation costs
 - Cons: the software can vanish or change at any time

- **Web application software**
 - Pros: easy to use, no installation costs
 - Cons: the software can vanish or change at any time
- **Dataverse Network Software**

- **Web application software**
 - Pros: easy to use, no installation costs
 - Cons: the software can vanish or change at any time
- **Dataverse Network Software**
 - Affero GPL License

- **Web application software**
 - Pros: easy to use, no installation costs
 - Cons: the software can vanish or change at any time
- **Dataverse Network Software**
 - Affero GPL License
 - Open source, public ownership

- **Web application software**
 - Pros: easy to use, no installation costs
 - Cons: the software can vanish or change at any time
- **Dataverse Network Software**
 - Affero GPL License
 - Open source, public ownership
 - If you don't like the new version: you can make a new one

- **Web application software**
 - Pros: easy to use, no installation costs
 - Cons: the software can vanish or change at any time
- **Dataverse Network Software**
 - Affero GPL License
 - Open source, public ownership
 - If you don't like the new version: you can make a new one
 - You own the software & the underlying code

- **Web application software**
 - Pros: easy to use, no installation costs
 - Cons: the software can vanish or change at any time
- **Dataverse Network Software**
 - Affero GPL License
 - Open source, public ownership
 - If you don't like the new version: you can make a new one
 - You own the software & the underlying code
 - The license guarantees that this will remain true in the future

- **Web application software**
 - Pros: easy to use, no installation costs
 - Cons: the software can vanish or change at any time
- **Dataverse Network Software**
 - Affero GPL License
 - Open source, public ownership
 - If you don't like the new version: you can make a new one
 - You own the software & the underlying code
 - The license guarantees that this will remain true in the future
- **Licensing data**

- **Web application software**
 - Pros: easy to use, no installation costs
 - Cons: the software can vanish or change at any time
- **Dataverse Network Software**
 - Affero GPL License
 - Open source, public ownership
 - If you don't like the new version: you can make a new one
 - You own the software & the underlying code
 - The license guarantees that this will remain true in the future
- **Licensing data**
 - DVN automates the IRB process; no lawyers necessary

- **Web application software**
 - Pros: easy to use, no installation costs
 - Cons: the software can vanish or change at any time
- **Dataverse Network Software**
 - Affero GPL License
 - Open source, public ownership
 - If you don't like the new version: you can make a new one
 - You own the software & the underlying code
 - The license guarantees that this will remain true in the future
- **Licensing data**
 - DVN automates the IRB process; no lawyers necessary
 - Data may be restricted in many ways, while metadata is available

<http://TheData.org>

Technology used in DVN Software

- Language: **Java Enterprise Edition 5** (with EJB3 and JSF) (team picked for JavaOne; Sun engineers regularly call for advice)

Technology used in DVN Software

- Language: **Java Enterprise Edition 5** (with EJB3 and JSF) (team picked for JavaOne; Sun engineers regularly call for advice)
- Application server: **GlassFish** (wrote press release on our project)

Technology used in DVN Software

- Language: **Java Enterprise Edition 5** (with EJB3 and JSF) (team picked for JavaOne; Sun engineers regularly call for advice)
- Application server: **GlassFish** (wrote press release on our project)
- Database: we use **PostgreSQL** (can substitute others)

Technology used in DVN Software

- Language: **Java Enterprise Edition 5** (with EJB3 and JSF) (team picked for JavaOne; Sun engineers regularly call for advice)
- Application server: **GlassFish** (wrote press release on our project)
- Database: we use **PostgreSQL** (can substitute others)
- Statistical computing: **R** and **Zelig**