Persistent Identifiers in Practice

Christian Zimmermann (Federal Reserve Bank of St. Louis)

Berlin 2012

The views expressed are those of individual authors and do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis, the Federal Reserve System, or the Board of Governors.

Why me?

Why me?

- Research Papers in Economics (RePEc)

Why me?

- Research Papers in Economics (RePEc)
- Federal Reserve Economic Data (FRED)

What is RePEc?

What is RePEc?

▶ Online bibliography for Economics, started in 1997 (1992)

What is RePEc?

- Online bibliography for Economics, started in 1997 (1992)
- Network of volunteers archivists

What is RePEc?

- ▶ Online bibliography for Economics, started in 1997 (1992)
- ▶ Network of volunteers archivists
- ▶ Metadata and retrieval protocols

What is RePEc?

- ▶ Online bibliography for Economics, started in 1997 (1992)
- ▶ Network of volunteers archivists
- ▶ Metadata and retrieval protocols
- ▶ Paper hosting at publishing institutions

What is RePEc?

- ▶ Online bibliography for Economics, started in 1997 (1992)
- ▶ Network of volunteers archivists
- ▶ Metadata and retrieval protocols
- ▶ Paper hosting at publishing institutions
- ▶ $\implies$ No central cost, everything free

What is RePEc?

- ▶ Online bibliography for Economics, started in 1997 (1992)
- ▶ Network of volunteers archivists
- ▶ Metadata and retrieval protocols
- ▶ Paper hosting at publishing institutions
- ▶ $\Longrightarrow$ No central cost, everything free
- ▶ 1,2 million items from 1500 journals and 3000 preprint series. 1400 archives

Organization

Organization

▶ Central list of pointers to metadata archives

Organization

- ▶ Central list of pointers to metadata archives
- ▶ Metadata archive following strict syntax

Organization

- ▶ Central list of pointers to metadata archives
- ▶ Metadata archive following strict syntax
- ▶ Automatical nightly retrieval of data

Organization

- ▶ Central list of pointers to metadata archives
- ▶ Metadata archive following strict syntax
- ▶ Automatical nightly retrieval of data
- ▶ "RePEc services" display metadata

Sample metadata template

Template-Type: ReDIF-Article 1.0
Author-Name: Andrew Beacon
Author-Name: Charlie Doe
Author-Email: doe@someuni.edu
Author-Workplace-Name: Some Big University
Author-Name: Engelbert Falafel
Author-Workplace-Name: Some Other University
Title: A Guide to Forecasting Agricultural Output by Following Weather Patterns in Neighboring Areas
Abstract: In this article, we survey the literature on agricultural output forecasting and suggest some avenues the
merit particular attention...
Classification-JEL: R00, Z0
Keywords: Forecasting, agricultural economics
Journal: Annals of Computational Economics
Pages: 1-58
Volume: 3
Issue: 1
Year: 1999
Month: February
File-URL: http://www.geekeconomics.com/articles/thefile.pdf
File-Format: application/pdf
File-URL: http://www.geekeconomics.com/articles/thefile.html
File-Format: text/html
Handle: RePEc:aaa:journl:v:3:y:1999:i:1:p:1-58

Metadata types

Metadata types

- archive

Metadata types

- ▶ archive
- ▶ series

Metadata types

- ► archive
- ► series
- ► working paper
- ► article
- ► book
- ► book chapter
- ► software component

Metadata types

- archive
- series
- working paper
- article
- book
- book chapter
- software component
- institution

Metadata types

- archive
- series
- working paper
- article
- book
- book chapter
- software component
- institution
- person

Metadata types

- ▶ archive
- ▶ series
- ▶ working paper
- ▶ article
- ▶ book
- ▶ book chapter
- ▶ software component
- ▶ institution
- ▶ person
- ▶ data?

Maintaining metadata

Maintaining metadata

- many different maintainers

Maintaining metadata

- many different maintainers
- varying motivations

Maintaining metadata

- ▶ many different maintainers
- ▶ varying motivations
- ▶ succession issues

Maintaining metadata

- ▶ many different maintainers
- ▶ varying motivations
- ▶ succession issues
- ▶ technical competence

Maintaining metadata

- many different maintainers
- varying motivations
- succession issues
- technical competence
- understanding of metadata

The Handle

The Handle

▶ Persistent identifier

The Handle

- ▶ Persistent identifier
- ▶ Used throughout RePEc services

The Handle

- ▶ Persistent identifier
- ▶ Used throughout RePEc services
- ▶ Unique

The Handle

- ▶ Persistent identifier
- ▶ Used throughout RePEc services
- ▶ Unique
- ▶ Permanent

The Handle

- ▶ Persistent identifier
- ▶ Used throughout RePEc services
- ▶ Unique
- ▶ Permanent
- ▶ Unique-Identifier-Never-Change-It

Why not persistent?

Why not persistent?

► Lack of understanding

Why not persistent?

- ► Lack of understanding
- ► Platform change

Why not persistent?

- ▶ Lack of understanding
- ▶ Platform change
- ▶ Recycling

Why not persistent?

- ▶ Lack of understanding
- ▶ Platform change
- ▶ Recycling
- ▶ No central authority

Introduction

Introduction

- Macroeconomic time series collection since 1993

Introduction

- ▶ Macroeconomic time series collection since 1993
- ▶ Database with 45,000 series

Introduction

- ► Macroeconomic time series collection since 1993
- ► Database with 45,000 series
- ► Graphing facility

Introduction

- ▶ Macroeconomic time series collection since 1993
- ▶ Database with 45,000 series
- ▶ Graphing facility
- ▶ Integrated in statistical packages

Introduction

- Macroeconomic time series collection since 1993
- Database with 45,000 series
- Graphing facility
- Integrated in statistical packages
- Apps, other tools

Introduction

- Macroeconomic time series collection since 1993
- Database with 45,000 series
- Graphing facility
- Integrated in statistical packages
- Apps, other tools
- ALFRED, GeoFRED

Metadata for data

Metadata for data

- United States: many statistical agencies (plus rest of world)

Metadata for data

- United States: many statistical agencies (plus rest of world)
- No central authority

Metadata for data

- United States: many statistical agencies (plus rest of world)
- No central authority
- No consensus metadata format, or data format

Metadata for data

- ▶ United States: many statistical agencies (plus rest of world)
- ▶ No central authority
- ▶ No consensus metadata format, or data format
- ▶ $\implies$ Just do it

The FRED approach

The FRED approach

- Set one's own standard

The FRED approach

- ▶ Set one's own standard
- ▶ Identifier schema in place for new data

The FRED approach

- ▶ Set one's own standard
- ▶ Identifier schema in place for new data
- ▶ Backward compatibility

The FRED approach

- ▶ Set one's own standard
- ▶ Identifier schema in place for new data
- ▶ Backward compatibility
- ▶ Very few persistence issues with sources

Links

- http://repec.org/
- http://research.stlouisfed.org/fred2/
- zimmermann@stlouisfed.org