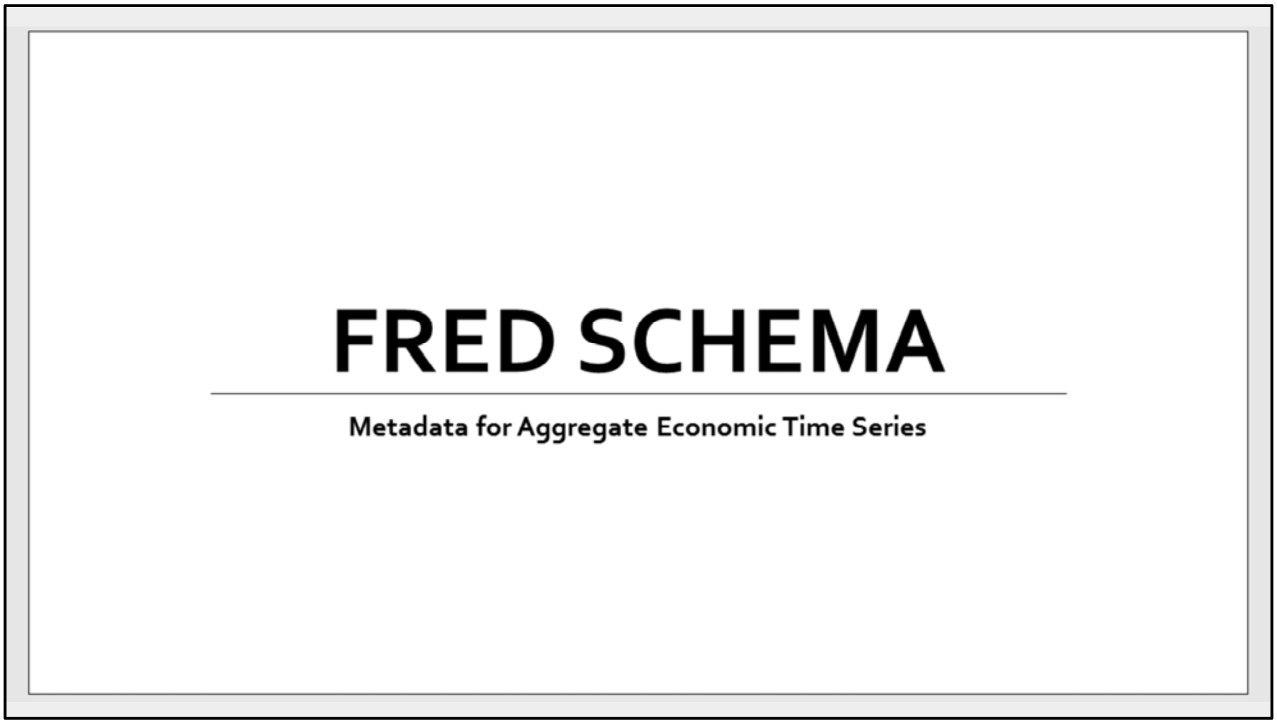
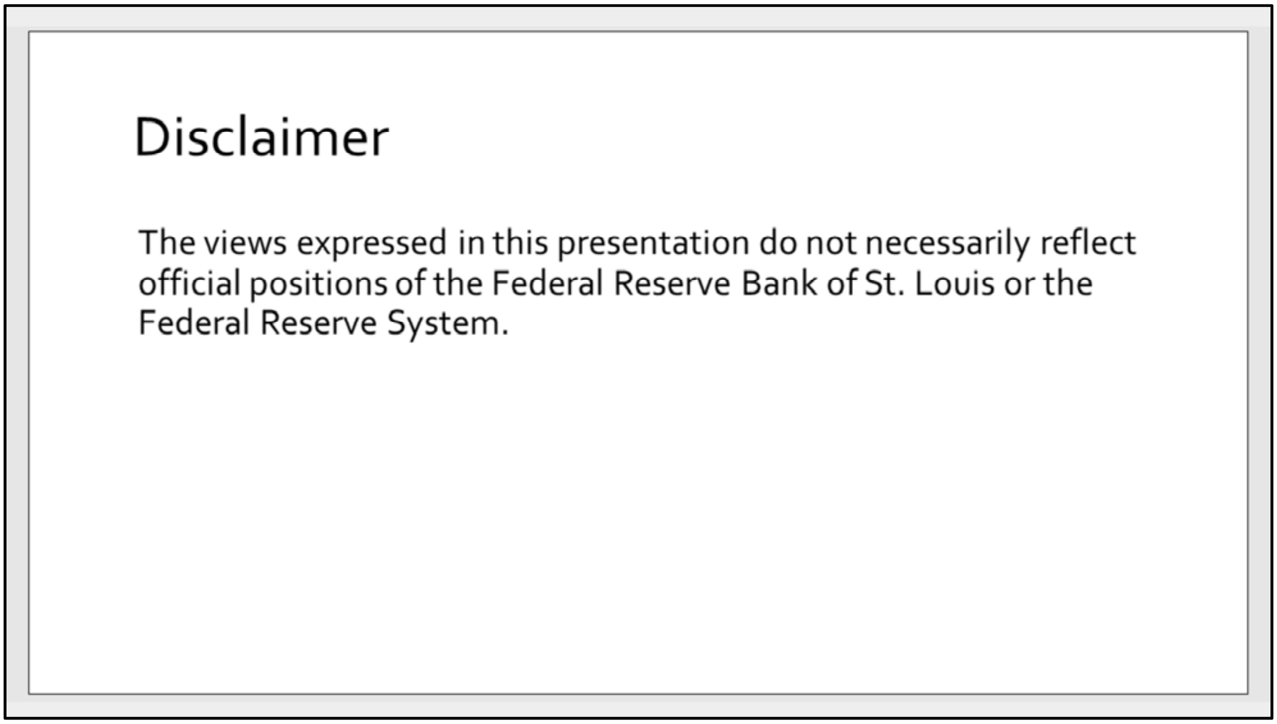


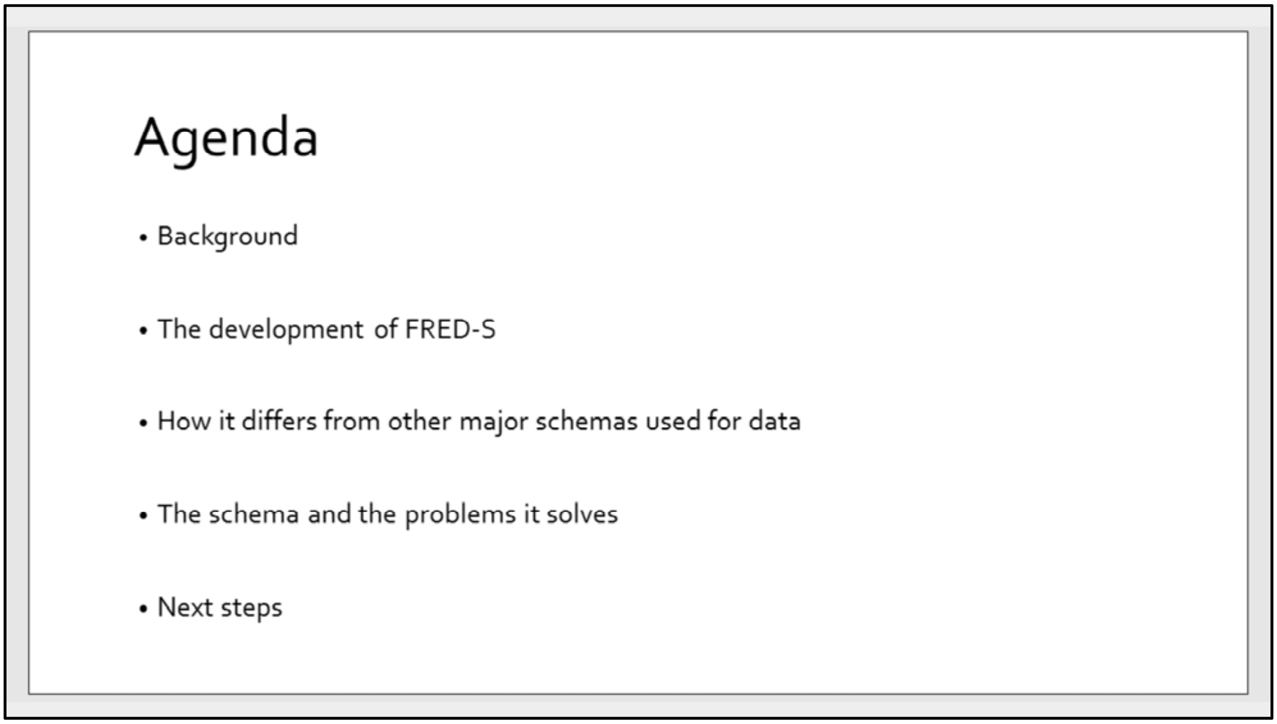
FRED SCHEMA

Metadata for Aggregate Economic Time Series



Disclaimer

The views expressed in this presentation do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis or the Federal Reserve System.

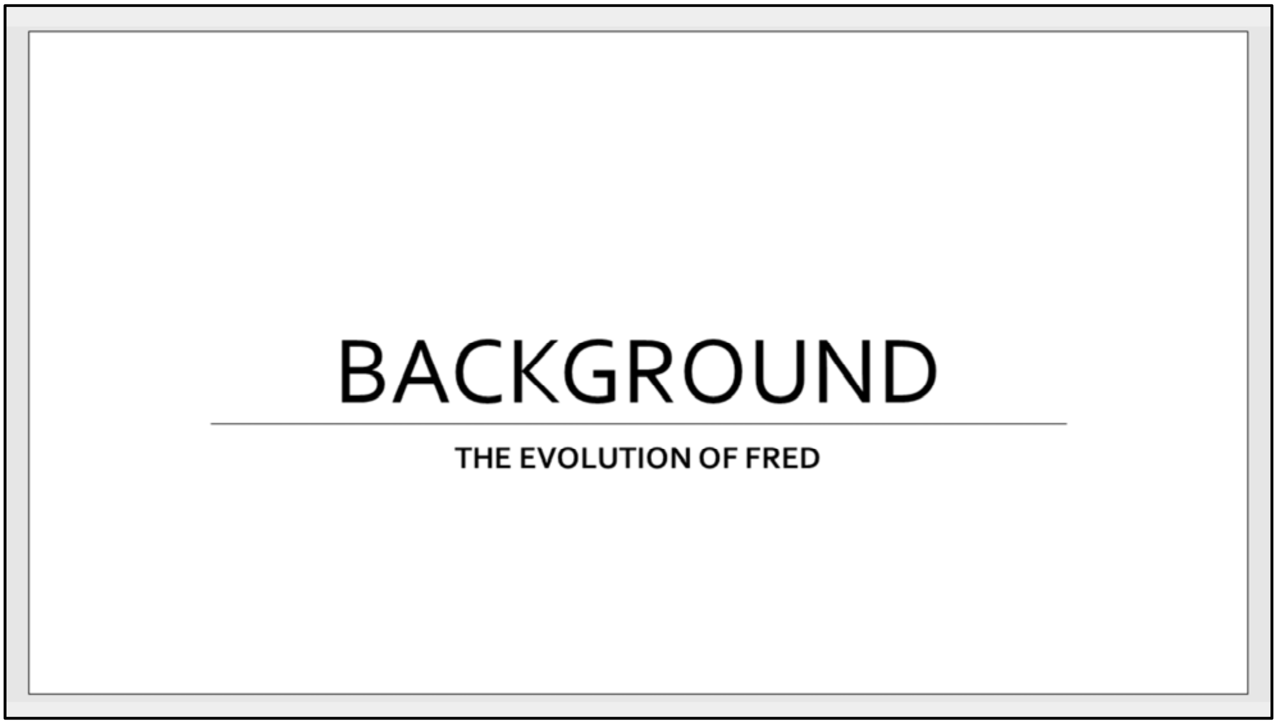


Agenda

- Background
- The development of FRED-S
- How it differs from other major schemas used for data
- The schema and the problems it solves
- Next steps

This morning I'm going to spend a little time explaining the background of this project
– why we decided to embark on this in the first place, and what our groundwork was
-- the process of developing FRED Schema
-- and why we thought it was really REALLY necessary to create a new schema when the metadata community is increasingly trying to streamline and combine
-- I'll then spend some time getting into the nitty-gritty of the schema and the specific problems that we hope to solve with our descriptions
-- and then I'll spend a minute or two talking about our plans and grand ideas, and leave a few minutes for questions at the end.

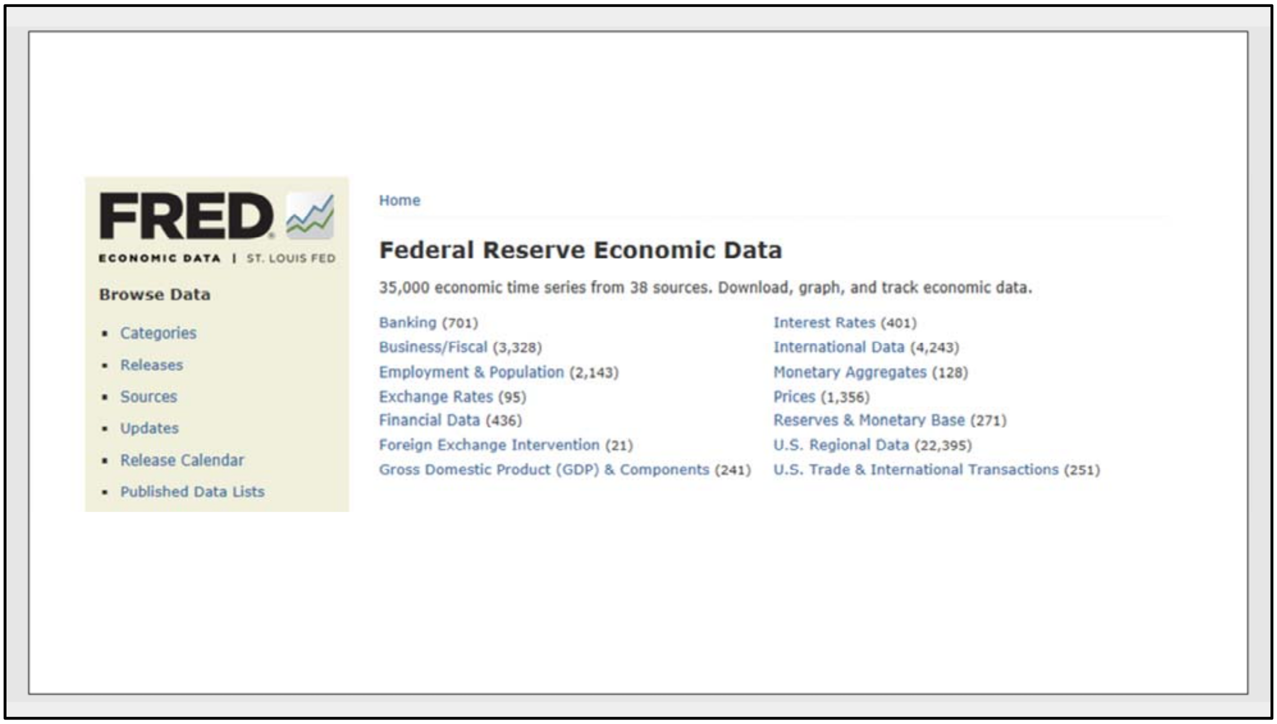
I have a terrible tendency to speed-present, and I've been working on this project for two years running, so please do interrupt me if I lose you along the way.



BACKGROUND

THE EVOLUTION OF FRED

Briefly, in case you don’t know, FRED is the St. Louis Fed’s free, publicly-available online database of economic (and economics-adjacent) data time series from national, international, public, and private sources.



FRED

ECONOMIC DATA | ST. LOUIS FED

Browse Data

- Categories
- Releases
- Sources
- Updates
- Release Calendar
- Published Data Lists

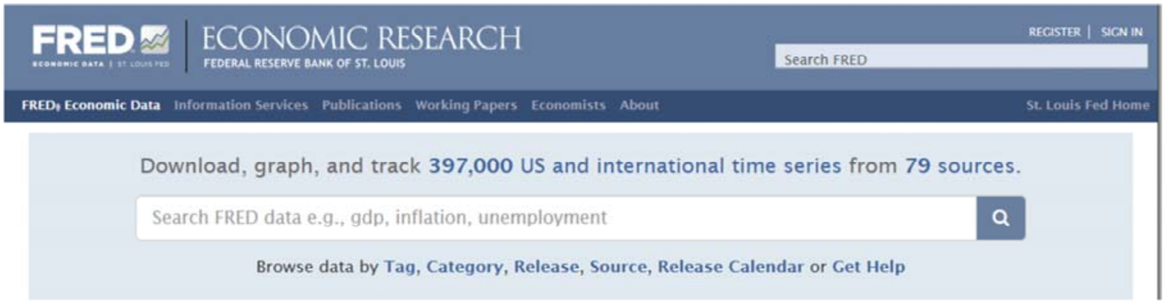
Home

Federal Reserve Economic Data

35,000 economic time series from 38 sources. Download, graph, and track economic data.

- Banking (701)
- Business/Fiscal (3,328)
- Employment & Population (2,143)
- Exchange Rates (95)
- Financial Data (436)
- Foreign Exchange Intervention (21)
- Gross Domestic Product (GDP) & Components (241)
- Interest Rates (401)
- International Data (4,243)
- Monetary Aggregates (128)
- Prices (1,356)
- Reserves & Monetary Base (271)
- U.S. Regional Data (22,395)
- U.S. Trade & International Transactions (251)

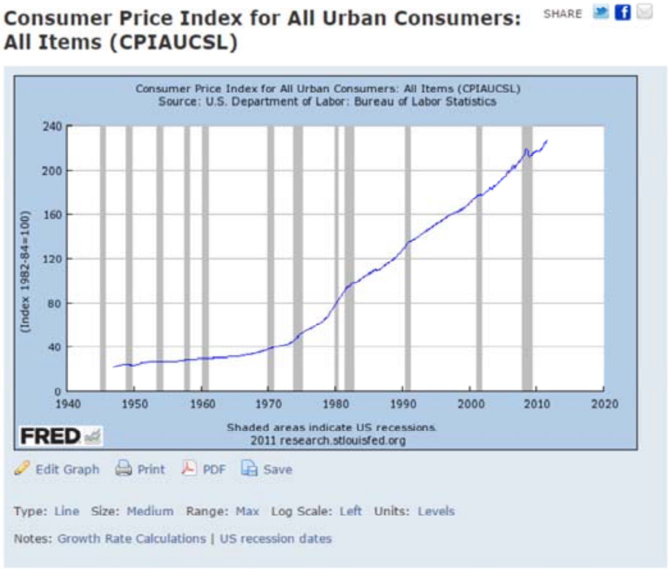
Five years ago, FRED had 35,000 data series and an average of about 150,000 visitors a month. By 2011, some version of FRED had been up and running for twenty years, from the bulletin board days to the early web to the FRED graphing tool, introduced in 2006 – which was also the year that we introduced the ability to download data to Excel.



Download, graph, and track 397,000 US and international time series from 79 sources.

Browse data by Tag, Category, Release, Source, Release Calendar or Get Help

Today, FRED has almost four hundred thousand series and an average of more than 600,000 visitors a month. Between 2011 and 2016 we also introduced an Excel add-in, and Android and iOS apps. That’s a lot of people accessing a lot of data in a lot of different ways, and for a lot of different purposes – and that’s been reflected in changes to the site itself.



Consumer Price Index for All Urban Consumers: All Items (CPIAUCSL)

SHARE
Consumer Price Index for All Urban Consumers: All Items (CPIAUCSL)
Source: U.S. Department of Labor: Bureau of Labor Statistics
240
200
160
120
80
40
0
(Index 1982-84=100)
1940
1950
1960
1970
1980
1990
2000
2010
2020
Shaded areas indicate US recessions.
2011 research.stlouisfed.org
FRED
Edit Graph
Print
PDF
Save

Type: Line Size: Medium Range: Max Log Scale: Left Units: Levels

Notes: Growth Rate Calculations | US recession dates

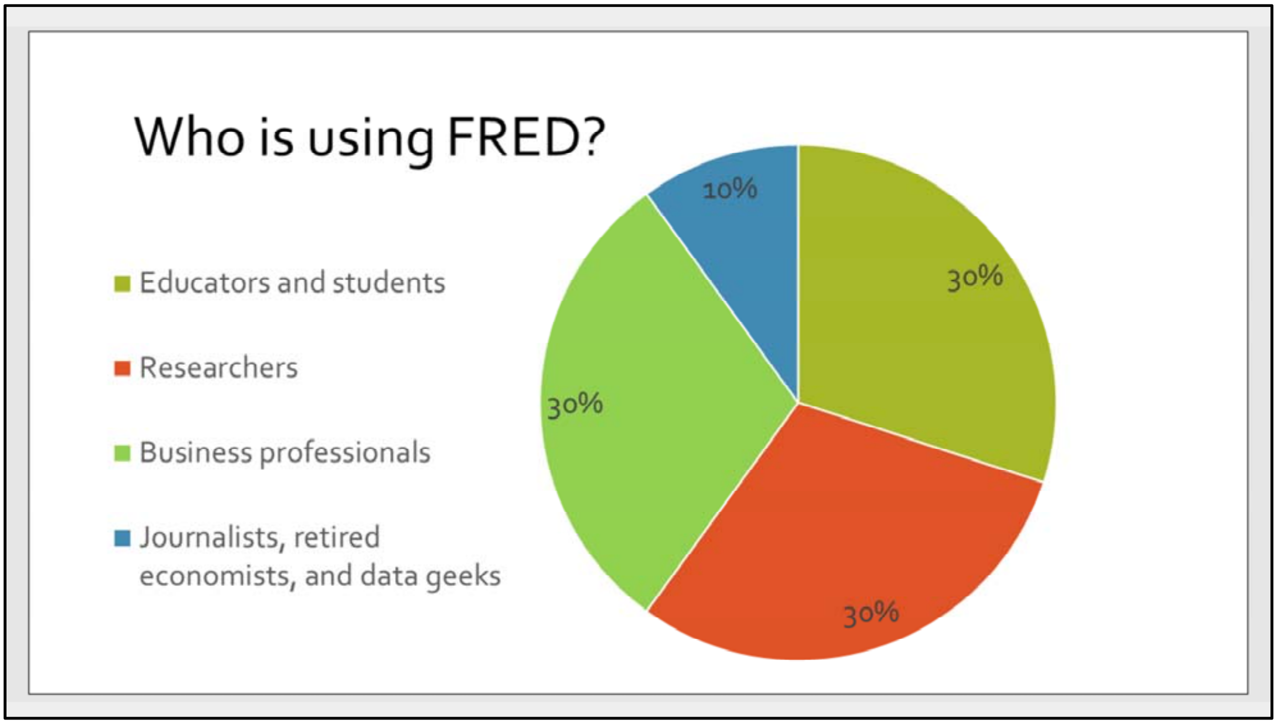
FRED series page for Consumer Price Index for All Urban Consumers, All Items, c. October 2011

- I know this looks ancient, but this is a screenshot of the series page from FRED as of late 2011.

- There's minimal metadata here – it tells you some details about the parameters of the graph itself (that it's a line, it's size medium, it shows the maximum range for this series, the number labeling is on the left, and the units are levels)

- On the actual page, below the area where I've grabbed this screenshot, there's some additional metadata – the series ID, the source (which is also below the title), the title of the release, the units (which match the axis label), the frequency, whether or not it's seasonally adjusted, the range of observations, and the last updated date. There are also links to the handbook of methods and the guide for understanding the CPI, from the source. That's it.

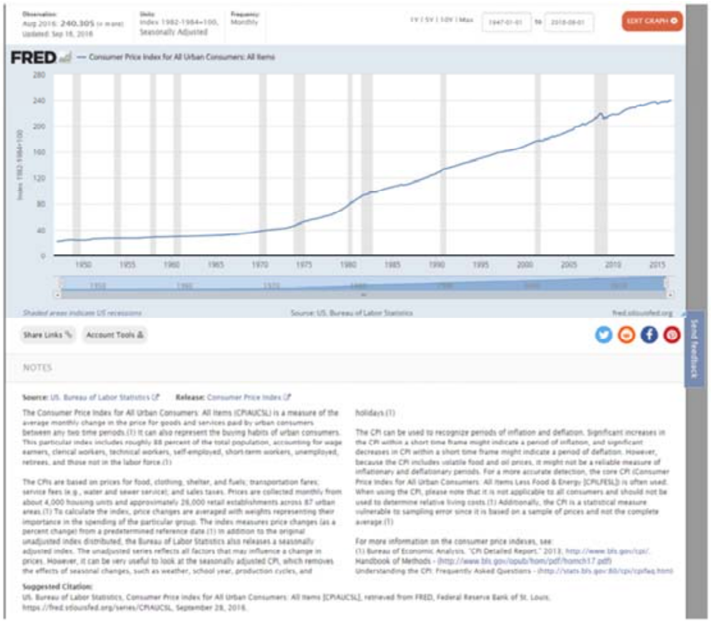
- This page makes a lot of assumptions about the knowledge and ability of its users – this is a page for users to make use of a convenient tool to view and work with data they're already familiar with. It doesn't teach, and it doesn't really explain itself in any way.



Who is using FRED?

Who uses FRED today? Those hundreds of thousands of monthly users are a mix: about a third of the audience is educators and their students, another third is researchers, another third is business professionals, and the last tenth are a mix of everyone else who needs or wants to use economic data – journalists, retirees, voters, data geeks – it's a broad variety.

We can literally have a first time economics student from high school and a Nobel Prize winning economist on the site at the same time – and one of the things we've found in doing user studies and getting feedback is that many of our users are less sophisticated data users than we might expect from their credentials – even those with a very deep knowledge within their focused subject expertise, may not know the idiosyncrasies of data series that may be relevant but are tangential to their research expertise.



Observation:
Aug 2016: 240.305 (+ more)
Updated: Sep 16, 2016

Units:
Index 1982-1984=100,
Seasonally Adjusted

Frequency:
Monthly

Shaded areas indicate US recessions

Source: US. Bureau of Labor Statistics

fred.stlouisfed.org

NOTES

Source: US. Bureau of Labor Statistics **Release:** Consumer Price Index

The Consumer Price Index for All Urban Consumers: All Items (CPIAUCSL) is a measure of the average monthly change in the price for goods and services paid by urban consumers between any two time periods.(1) It can also represent the buying habits of urban consumers. This particular index includes roughly 88 percent of the total population, accounting for wage earners, clerical workers, technical workers, self-employed, short-term workers, unemployed, retirees, and those not in the labor force.(1)

The CPIs are based on prices for food, clothing, shelter, and fuels; transportation fares; service fees (e.g., water and sewer service); and sales taxes. Prices are collected monthly from about 4,000 housing units and approximately 26,000 retail establishments across 87 urban areas.(1) To calculate the index, price changes are averaged with weights representing their importance in the spending of the particular group. The index measures price changes (as a percent change) from a predetermined reference date.(1) In addition to the original unadjusted index distributed, the Bureau of Labor Statistics also releases a seasonally adjusted index. The unadjusted series reflects all factors that may influence a change in prices. However, it can be very useful to look at the seasonally adjusted CPI, which removes the effects of seasonal changes, such as weather, school year, production cycles, and holidays.(1)

The CPI can be used to recognize periods of inflation and deflation. Significant increases in the CPI within a short time frame might indicate a period of inflation, and significant decreases in CPI within a short time frame might indicate a period of deflation. However, because the CPI includes volatile food and oil prices, it might not be a reliable measure of inflationary and deflationary periods. For a more accurate detection, the core CPI (Consumer Price Index for All Urban Consumers: All Items Less Food & Energy [CPILFESL]) is often used. When using the CPI, please note that it is not applicable to all consumers and should not be used to determine relative living costs.(1) Additionally, the CPI is a statistical measure vulnerable to sampling error since it is based on a sample of prices and not the complete average.(1)

For more information on the consumer price indexes, see:
(1) Bureau of Economic Analysis. "CPI Detailed Report." 2013. http://www.bls.gov/cpi/.
Handbook of Methods - (http://www.bls.gov/opub/hom/pdf/homch17.pdf)
Understanding the CPI: Frequently Asked Questions - (http://stats.bls.gov:80/cpi/cpifaq.htm)

Suggested Citation:
US. Bureau of Labor Statistics, Consumer Price Index for All Urban Consumers: All Items [CPIAUCSL], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/CPIAUCSL, September 28, 2016.

FRED series page for Consumer Price Index for All Urban Consumers, All Items, c. October 2016

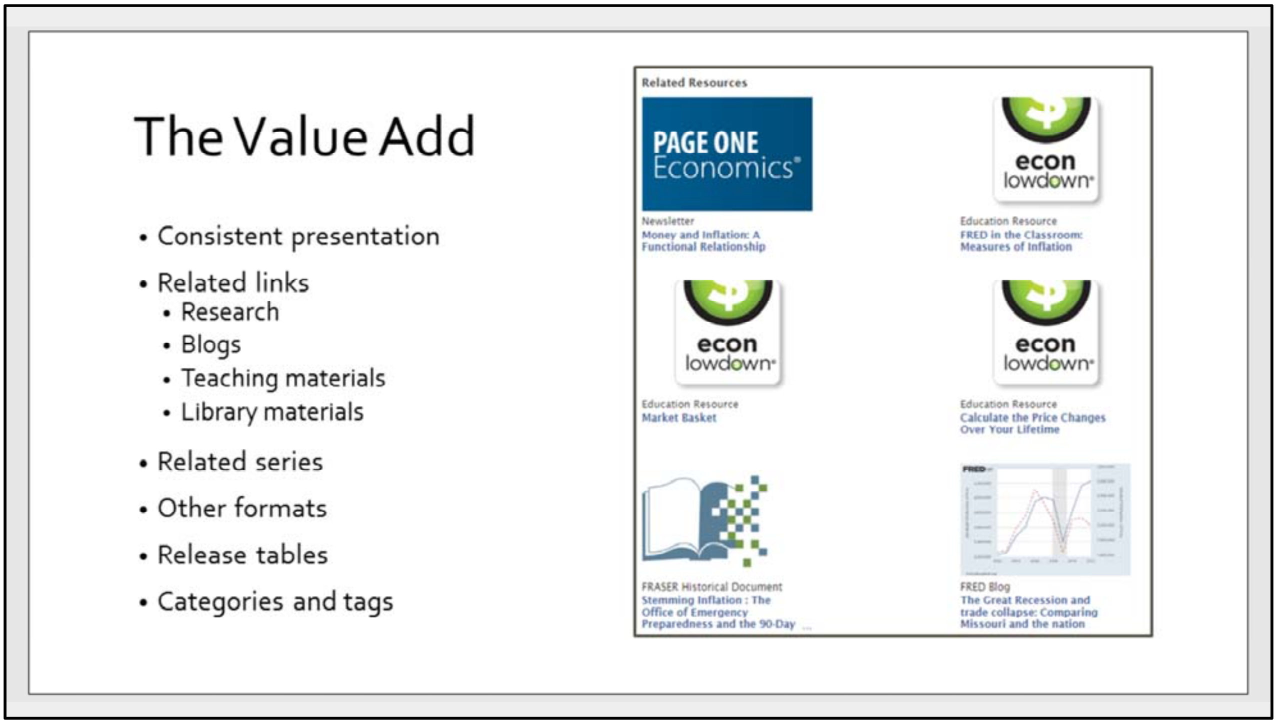
So in trying to address the needs of that wider – and widening audience, FRED has adapted to provide more and clearer contextual information

I know that this is probably too small to read, but that's ok –I really just want to give you an illustration of the volume of how **much** more information is available today

You've got a bigger graph with more bells and whistles (that's a date slider at the bottom). You've got metadata at the top, with observations, units, and frequency. You've got zoom-to-year buttons and action buttons for editing and sharing. But for me that's not the biggest thing – look at all the text!

First you've got links back to the source and the specific release. (Those were links to FRED pages in the earlier iteration)

Then you have ALL THIS TEXT. These are notes that explain briefly what CPI is, how it's calculated, and what you can use it for. The same links from the 2011 version are here again, but there's now a line that tells you what those links are and where they go. Finally, you have a suggested citation – because we understand that students and researchers are coming to FRED to find and use this data, not just view it.



The Value Add

- Consistent presentation
- Related links
  - Research
  - Blogs
  - Teaching materials
  - Library materials
- Related series
- Other formats
- Release tables
- Categories and tags

Related Resources

Newsletter
Money and Inflation: A Functional Relationship

Education Resource
FRED in the Classroom: Measures of Inflation

Education Resource
Market Basket

Education Resource
Calculate the Price Changes Over Your Lifetime

FRASER Historical Document
Stemming Inflation : The Office of Emergency Preparedness and the 90-Day ...

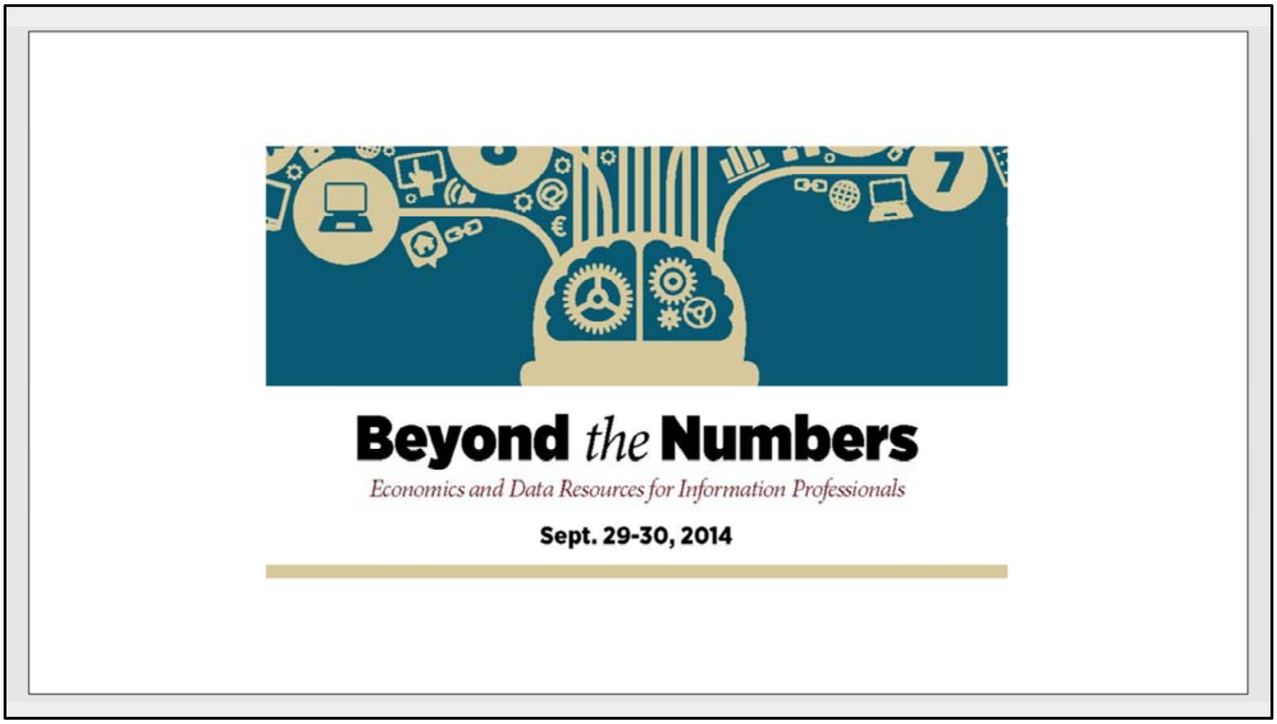
FRED Blog
The Great Recession and trade collapse: Comparing Missouri and the nation

In addition to sometimes lengthy explanatory notes, FRED series now offer

- related content links to other pages within the St. Louis Research domain, like blog posts, research papers, lessons, and digital copies of the original data releases on FRASER
- links to other formats (seasonally adjusted vs. not seasonally adjusted, millions instead of billions, etc.)
- links to FRED-formatted release tables, where those have been created
- browseable categories and tags

Nearly all of that context-giving content has to be assigned manually. That's a huge amount of work to do right – and the FRED team does it with just a team of five. We have a fundamental scalability problem, but it's built on an important commitment to the education of our users.

So this project starts with "we know we have a problem, we have a general sense of what the problem is, and we're pretty sure metadata might solve some of it"



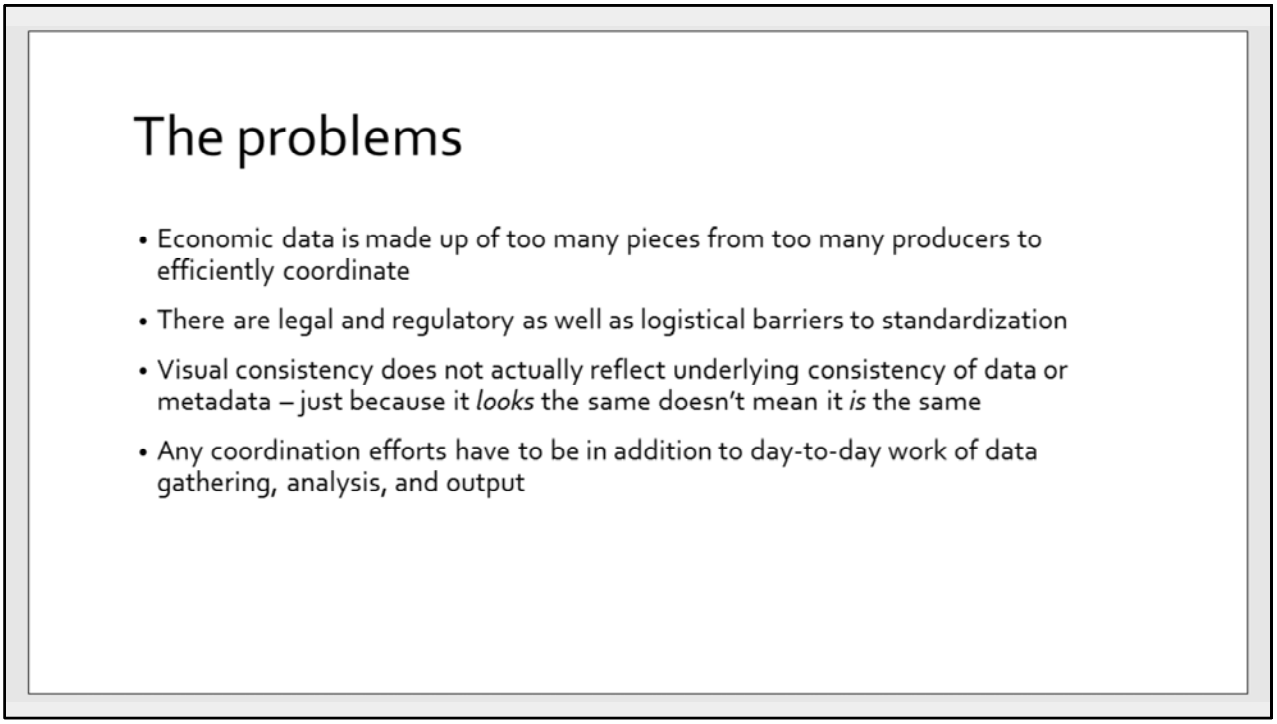
Beyond *the* Numbers

Economics and Data Resources for Information Professionals

Sept. 29-30, 2014

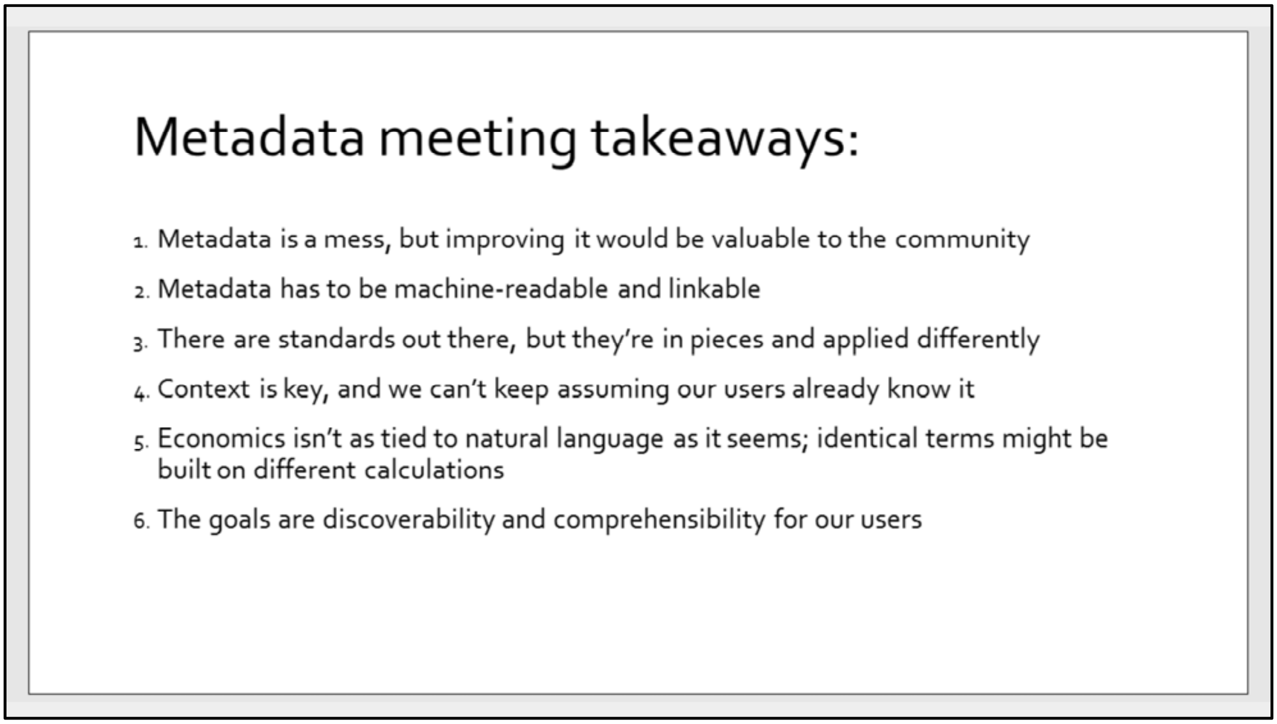
So despite the size of our lean, mean data team, FRED doesn’t work alone – whenever we can, we work **with** data producers to get their data into FRED in a way that’s as easy as possible for all parties -- but obviously there are a lot of idiosyncrasies when trying to translate from one information structure into another. When we held the first Beyond the Numbers, we realized that we had the opportunity to get a number of stakeholders in a room and start a conversation about what, if any, could be done to make the task easier and find common ground between institutions.

During that conference we met with representatives from Federal Reserve Banks, from NGOs, and from national and international data providers and tried to find common ground on the metadata issue.



The problems

- Economic data is made up of too many pieces from too many producers to efficiently coordinate
- There are legal and regulatory as well as logistical barriers to standardization
- Visual consistency does not actually reflect underlying consistency of data or metadata – just because it *looks* the same doesn't mean it *is* the same
- Any coordination efforts have to be in addition to day-to-day work of data gathering, analysis, and output



Metadata meeting takeaways:

1. Metadata is a mess, but improving it would be valuable to the community
2. Metadata has to be machine-readable and linkable
3. There are standards out there, but they're in pieces and applied differently
4. Context is key, and we can't keep assuming our users already know it
5. Economics isn't as tied to natural language as it seems; identical terms might be built on different calculations
6. The goals are discoverability and comprehensibility for our users

That meeting started a great discussion, and gave us some general organizing principles to work from as we tried to figure out our next steps

Number 1: Metadata is a mess, but improving it would be valuable to the community

- we have to start somewhere. Unlike in the EU where there's some ability to do top-down imposition of standards, in the US and internationally we have to patch things together – but it's worth it for our own uses and for our users' needs

Number 2: Metadata has to be machine-readable and linkable
- Human-readable metadata, like a 19th century catalog card, is great, but to make our data findable in the age of Google requires machine-readable metadata. Taking advantage of linked data protocols whenever possible should be our goal. About 60% of FRED's traffic, for instance, comes straight from Google directly to the series page – if we have insufficient context, we run the risk that either they'll end up on the wrong page because Google doesn't understand what's on the page, or we'll get bounce because users don't understand why Google dropped them on that page.

Number 3: There are standards out there, but they're in pieces and applied differently

- We don't want to reinvent the wheel (or gross domestic product) – we want to declare use of standards, not create them. By linking to standards and definitions established by consortiums and working groups, as well as by data producers themselves, we can save ourselves a lot of time and effort and our users a lot of headaches. And that's not just for FRED –guidance for where it's most helpful to link to a standard can be useful for any data site

Number 4: Context is key, and we can't keep assuming our users already know what it is

- Economic data used to be predominantly by and for economists. Journalists and non-economics students and high school teachers didn't need to know the intricacies of data updates and revisions – but if everyone and their mother is using data for all kinds of innovative uses, they need to be able to quickly find the resources to answer questions about what they're looking at.

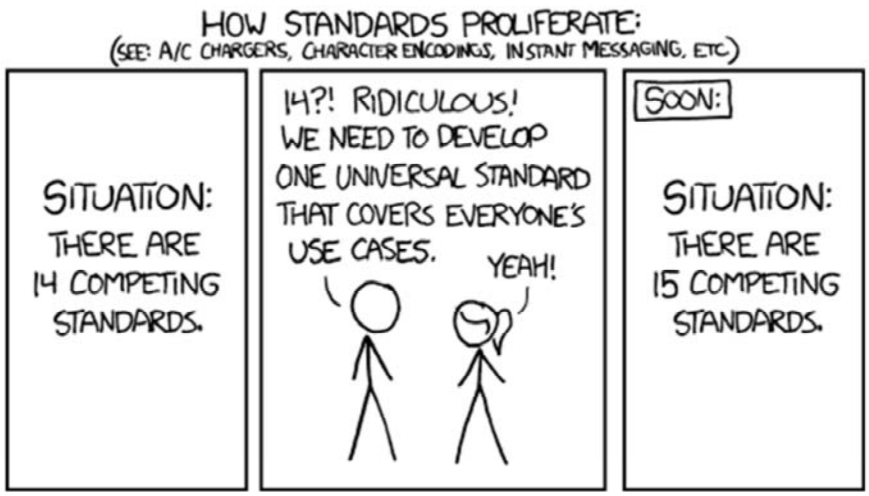
Number 5: Economics isn't as tied to natural language as it seems; identical terms might be built on different calculations

- Is it GDP or GDP? While theoretically the expenditures and income approaches to the calculation get you the same number, in practice they don't. The more we expand our user base, the more language can fail ordinary non-economists

Number 6: The goals are discoverability and comprehensibility for our users
- The most sophisticated metadata framework in the world is useless to someone who can't understand it or use it to navigate. It doesn't mean it's useless – schemas like DDI and SDMX are incredibly powerful for their intended uses – but a Phillips-head screwdriver makes a bad hammer.

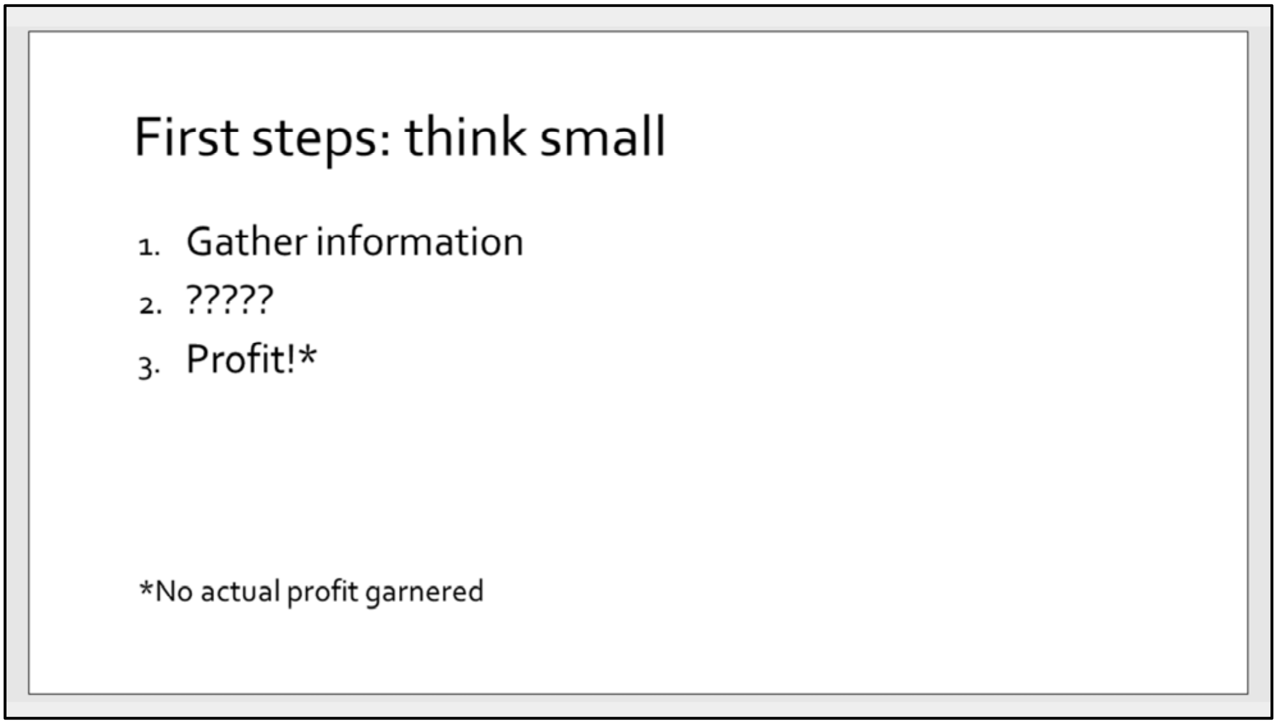
THE ACCIDENTAL SCHEMA

So let’s talk a little bit about how the FRED Schema was actually developed


HOW STANDARDS PROLIFERATE:
(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC)
SITUATION: THERE ARE 14 COMPETING STANDARDS.
14?! RIDICULOUS! WE NEED TO DEVELOP ONE UNIVERSAL STANDARD THAT COVERS EVERYONE'S USE CASES.
YEAH!
SOON:
SITUATION: THERE ARE 15 COMPETING STANDARDS.

🅭 https://xkcd.com/927/

At the earliest stages, we thought we were going to be creating detailed crosswalks or starting a data dictionary or information site. We already knew that competition between standards was a problem, and we wanted to fix -- not contribute to -- the confusion. Since this entire talk is about the metadata schema we created, I'm sure you can tell just how successful we were in that.



First steps: think small

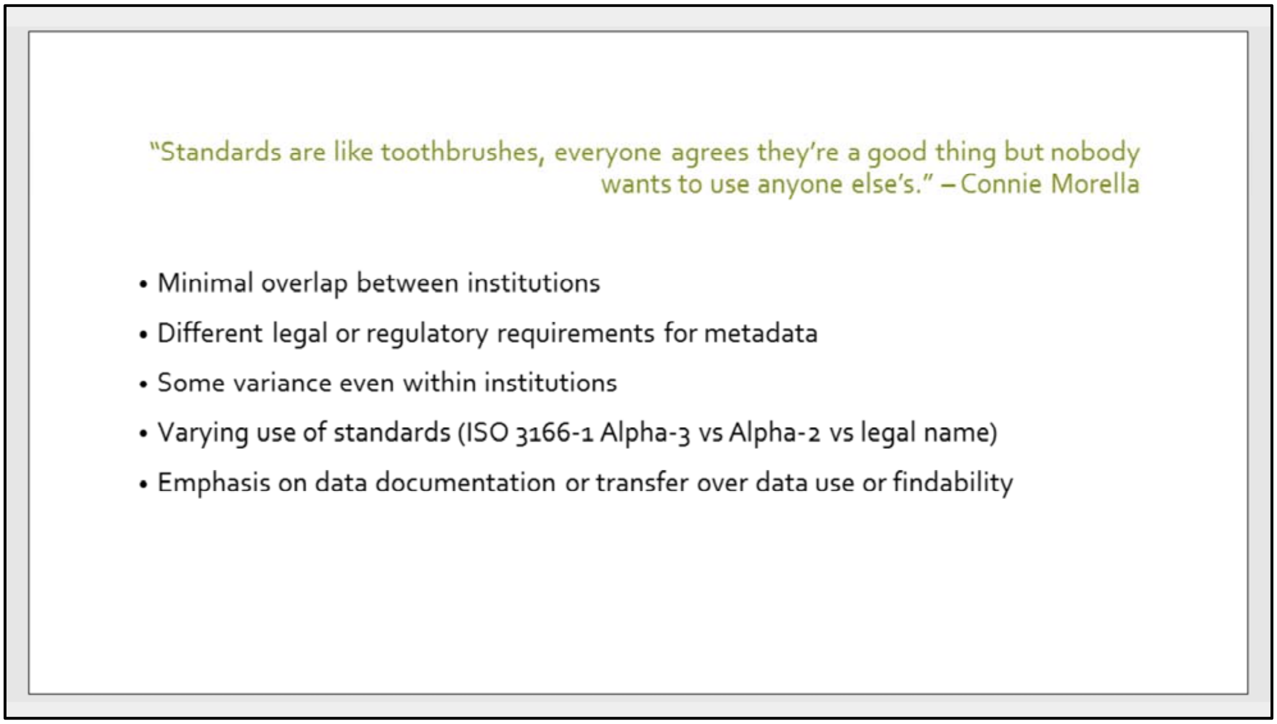
1. Gather information
2. ?????
3. Profit!*

*No actual profit garnered

So! Once we'd left that meeting at Beyond the Numbers with a solid commitment to at least explore what our options were and what the low-hanging fruit of the metadata vineyard might be, we embarked on two information-gathering projects:

First, we looked hard at what FRED was already capturing in its metadata, and what big chunks of contextual information were necessary to identify and describe a series. We wanted to get a sense of what metadata could and **should** be collected or assigned for time series metadata

Once we established our baseline, we created a survey and sent it out to the participants of that first meeting and some other interested organizations in the banking and economics domain to try to get a sense of what metadata standards or schemas they were using, and how. We wanted to see how their usual practices could be adapted to address FRED's core metadata needs, and where the overlap was between their metadata implementations.



"Standards are like toothbrushes, everyone agrees they're a good thing but nobody wants to use anyone else's." – Connie Morella

- Minimal overlap between institutions
- Different legal or regulatory requirements for metadata
- Some variance even within institutions
- Varying use of standards (ISO 3166-1 Alpha-3 vs Alpha-2 vs legal name)
- Emphasis on data documentation or transfer over data use or findability

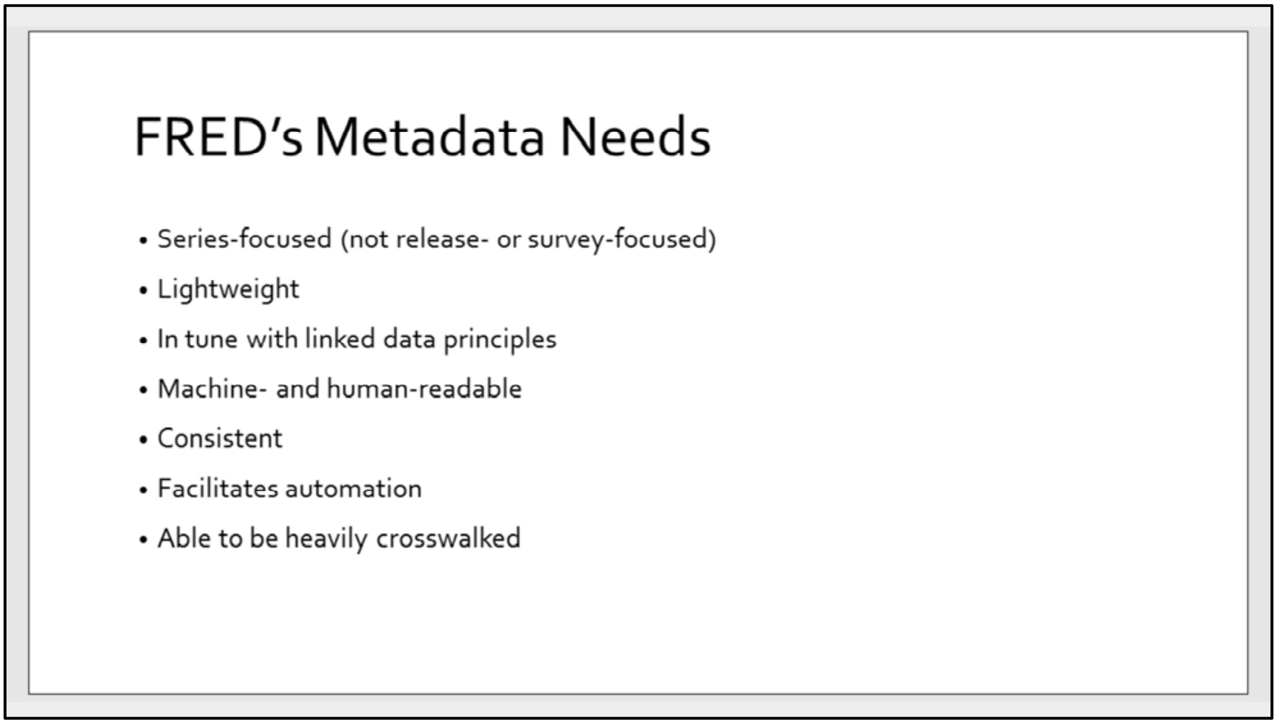
We surveyed about a half a dozen major economic data providers and found a lot of conceptual overlap without a lot of practical overlap.

There was minimal overlap between institutions
Different legal or regulatory requirements for metadata
Some variance even within institutions – if you saw Dan Gillman's talk about the metadata projects at the BLS you are aware of one example
There's varying use of standards (ISO 3166-1 Alpha-3 vs Alpha-2 vs legal name)
And a real emphasis on data documentation or transfer over data use or findability

Just a caveat – it's understandable that these organizations are interested in data transfer first, and use a metadata framework that makes their job easier. What we determined from the survey, however, is that we weren't going to find any ripe low-hanging fruit or any branches we could just refashion into a framework for the functions FRED needs to fulfil.

(I promise I'll stop torturing that metaphor now)

So what are FRED's metadata needs?



FRED's Metadata Needs

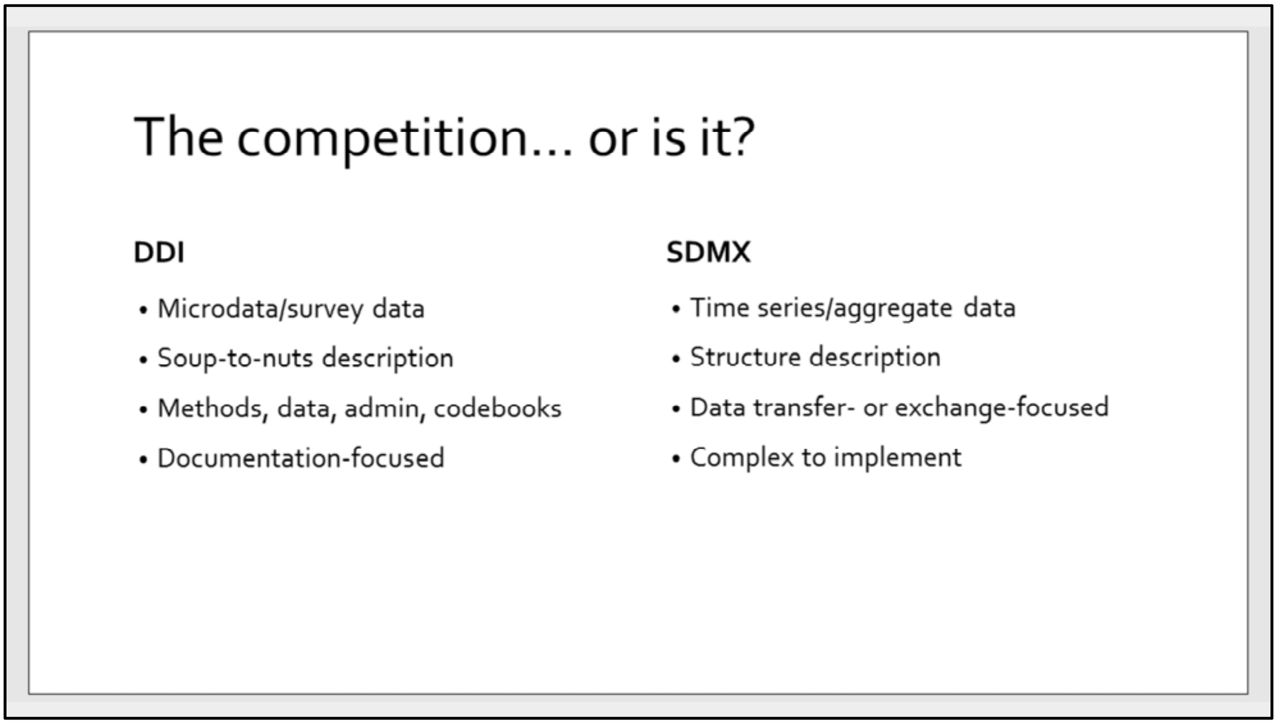
- Series-focused (not release- or survey-focused)
- Lightweight
- In tune with linked data principles
- Machine- and human-readable
- Consistent
- Facilitates automation
- Able to be heavily crosswalked

Currently, FRED series – the series description, not the individual observations and their metadata -- are built on essentially five metadata fields: id, unit, title, frequency, and seasonality, plus a free-text notes field, category, and tags. There has historically been very little consistency in the application of those elements.

We decided that if we were going to spend the time and energy doing a metadata overhaul – and another team I'm on had just recently finished one for FRASER, our digital library – we knew it needed to have some specific capabilities:

First, it needed to be series-focused (not release- or survey-focused)
It had to be lightweight but powerful, in tune with linked data principles, and machine- and human-readable.
It needed to be consistent but also to facilitate automation, and it needed to be something that could be heavily crosswalked.

FRED is just one player in the data game. We may have a great product, but we're not in the business of dictating practice for our partners. We wanted something that would serve our needs and provide opportunities for data providers whose content we aggregate to help us choose the right metadata for their data.



The competition… or is it?

DDI

- Microdata/survey data
- Soup-to-nuts description
- Methods, data, admin, codebooks
- Documentation-focused

SDMX

- Time series/aggregate data
- Structure description
- Data transfer- or exchange-focused
- Complex to implement

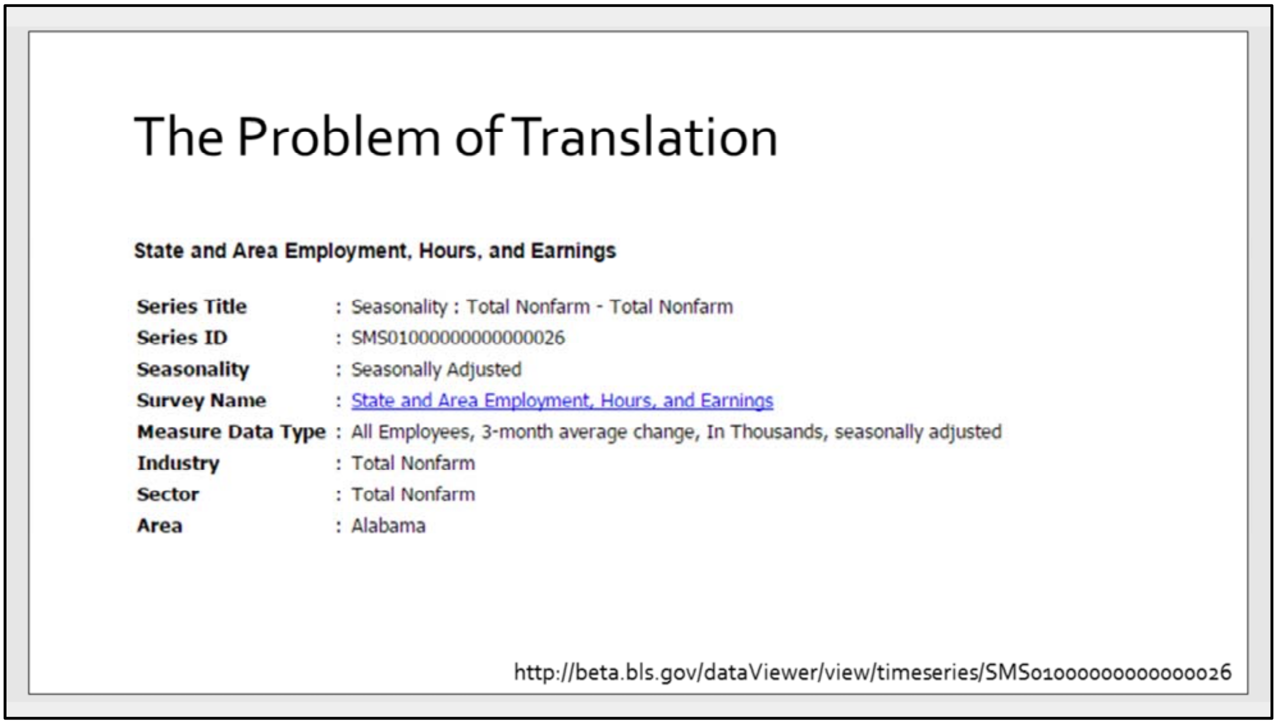
So let’s talk about those other metadata schemas

The two biggest players in the data space, at least in economics, seem to be DDI and SDMX

Those of you who are familiar with DDI, the data documentation initiative, know that it’s an incredibly powerful and flexible schema for describing statistical and social science data. When they say data, however, the DDI community generally means microdata or survey data - that must be transformed or compiled to become time series data.

SDMX is the big dog in economic time series data, which I know because everyone I talked to who uses it said “Well obviously you should be using SDMX”. SDMX is an international standard that streamlines the process of exchange of statistical data and its metadata. SDMX is very powerful, but its primary use case isn’t discoverability by novice or student users who may not already understand what they’re looking at. Bear in mind that FRED also already has a strong, highly-automated process for actual data transfer. My personal impression is that using SDMX without its data transfer functionality is sort of like using the jaws of life to remove a splinter: wildly overkill and probably also ineffective.

The other schema often used for data description is Dublin Core, which provides only the most basic bibliographic description. While it’s flexible, I’ve found that there’s so much extension required to use it for a specific purpose that it ends up being far less of a real “standard” standard than is needed.



The Problem of Translation

State and Area Employment, Hours, and Earnings

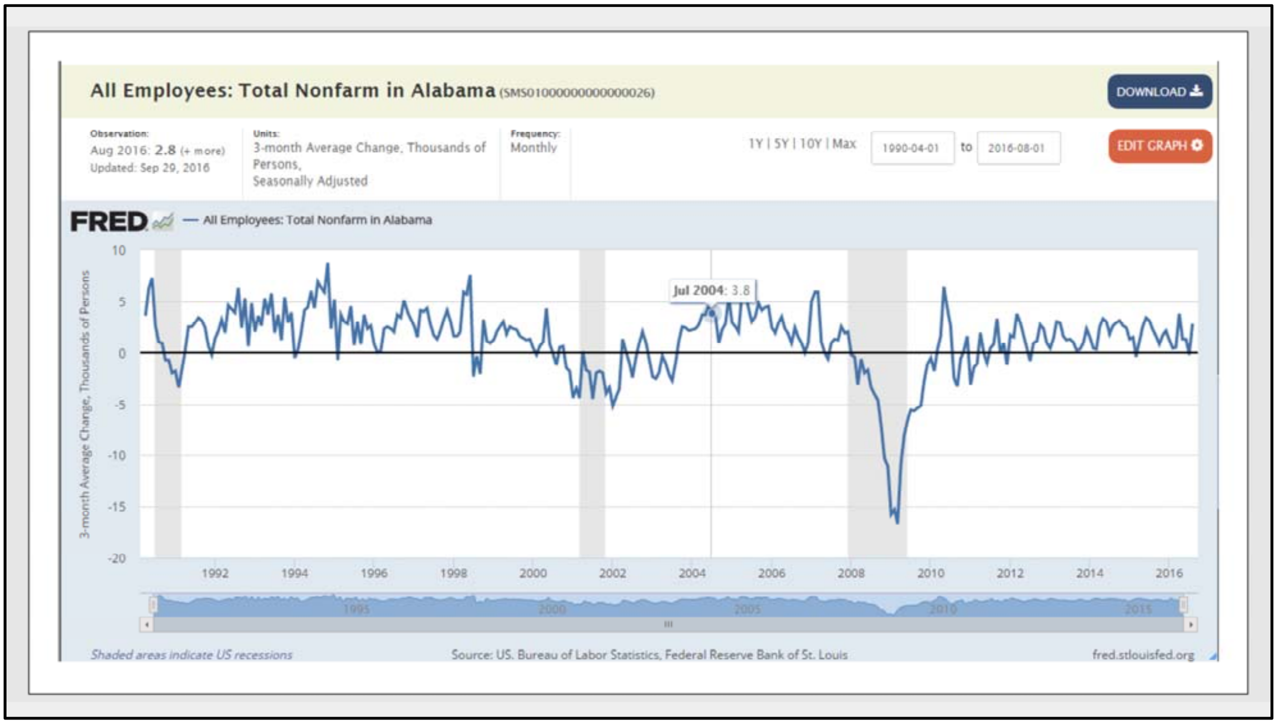
Series Title	: Seasonality : Total Nonfarm - Total Nonfarm
Series ID	: SMS01000000000000026
Seasonality	: Seasonally Adjusted
Survey Name	: State and Area Employment, Hours, and Earnings
Measure Data Type	: All Employees, 3-month average change, In Thousands, seasonally adjusted
Industry	: Total Nonfarm
Sector	: Total Nonfarm
Area	: Alabama

http://beta.bls.gov/dataViewer/view/timeseries/SMS01000000000000026

FRED has to do a lot of translation between internal or purpose-built metadata frameworks and something more general that can be used for aggregation.

This is actually the series that started FRED schema. This is a BLS state employment series that in July of last year we were trying to figure out how to pull into FRED in a way that made sense to our users. Here we've got a producer-provided title that doesn't stand alone, an industry and sector that aren't currently captured by FRED as separate classifications, and a "Measure Data Type" or unit that we had to actually do the math on to try to figure out the order of operations: Is it the average change, or the change in average? Is it seasonally adjusted before or after the averaging?

When you put that into FRED you get …



All Employees: Total Nonfarm in Alabama (SMS01000000000000026)

Observation:
Aug 2016: 2.8 (+ more)
Updated: Sep 29, 2016

Units:
3-month Average Change, Thousands of Persons,
Seasonally Adjusted

Frequency:
Monthly

1Y | 5Y | 10Y | Max 1990-04-01 to 2016-08-01

Shaded areas indicate US recessions

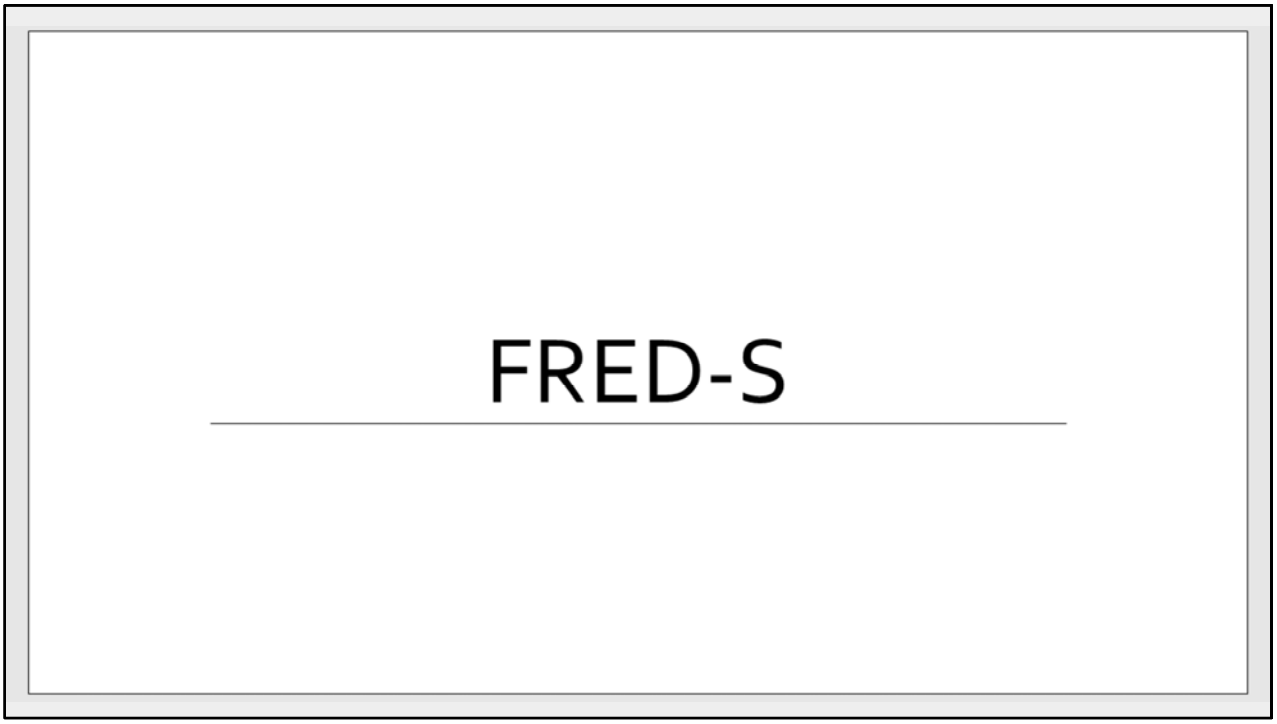
Source: US. Bureau of Labor Statistics, Federal Reserve Bank of St. Louis

fred.stlouisfed.org

A Title of “All Employees: Total Nonfarm in Alabama”
A Unit of “3-month Average Change, Thousands of Persons”

Seasonality stays the same, observations stay the same, but we’ve declared the frequency too.

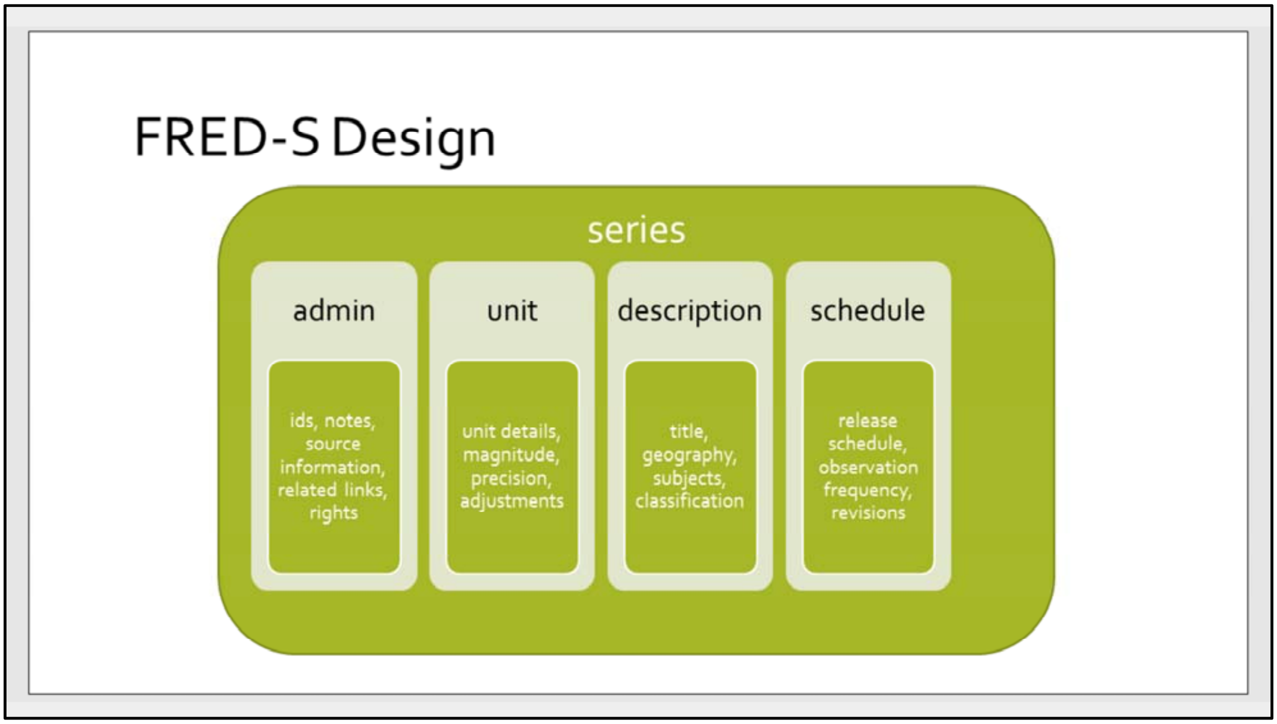
Trying to figure out the title and units on this series is what caused us to start thinking about the metadata modularly and realizing that we could do a lot more with a holistic series-level schema rather than a cobbling-together of elements from other metadata schemas.



FRED-S

When developing our schema, we kept one caveat in mind: we may be the only people who ever use this framework, so it needs to be built to work with, not compete with, other schemas.

FRED-S is designed to bring together information encoded in lots of places and establish links. It will never replace a powerful comprehensive schema like DDI or SDMX, but our hope is that it can declare the few key pieces of a time series and allow for better automation of linkages, with the goal of informing and educating a broad audience. If we do it right, a FRED-S implementation could sit on top of one of those other schemas like a particularly elegant hat.



FRED-S Design

series

admin

ids, notes, source information, related links, rights

unit

unit details, magnitude, precision, adjustments

description

title, geography, subjects, classification

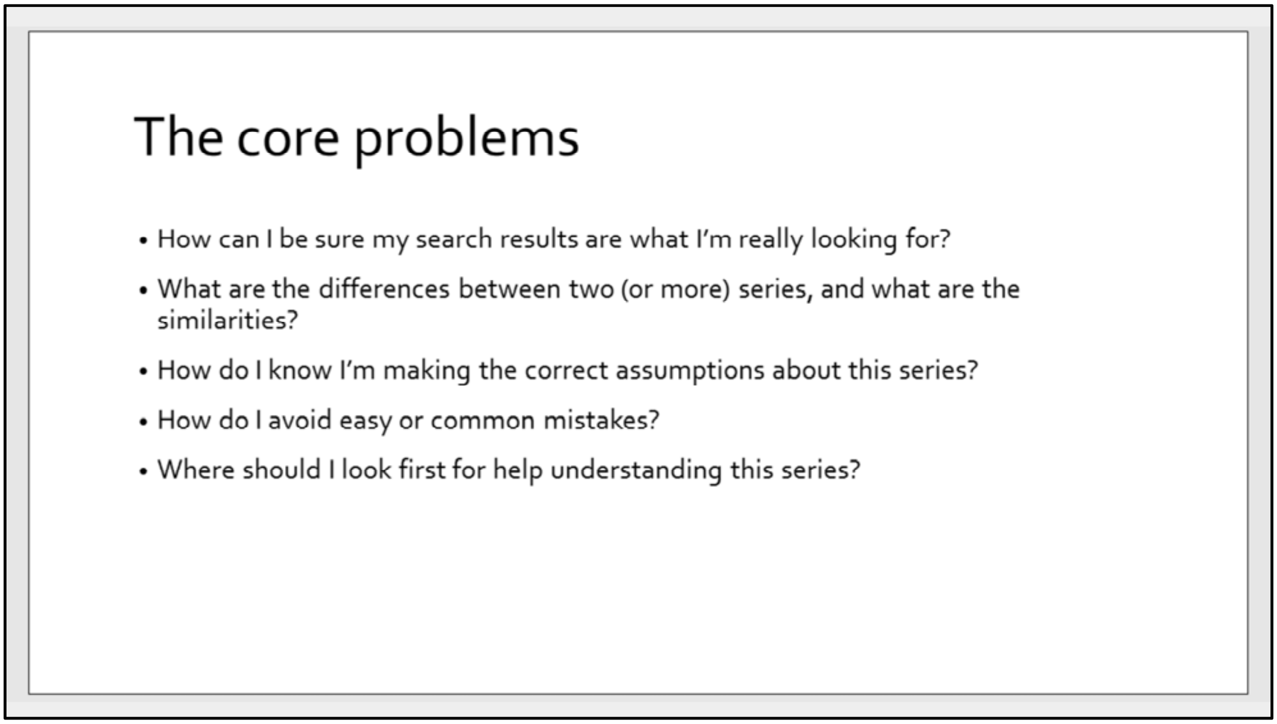
schedule

release schedule, observation frequency, revisions

FRED-S attempts to create a kind of catalog record for an individual time series:

- where does it come from and what’s it related to?
- how is the series measured
- what‘s the topic or concept that the series shows; and
- when does the data get released or updated?

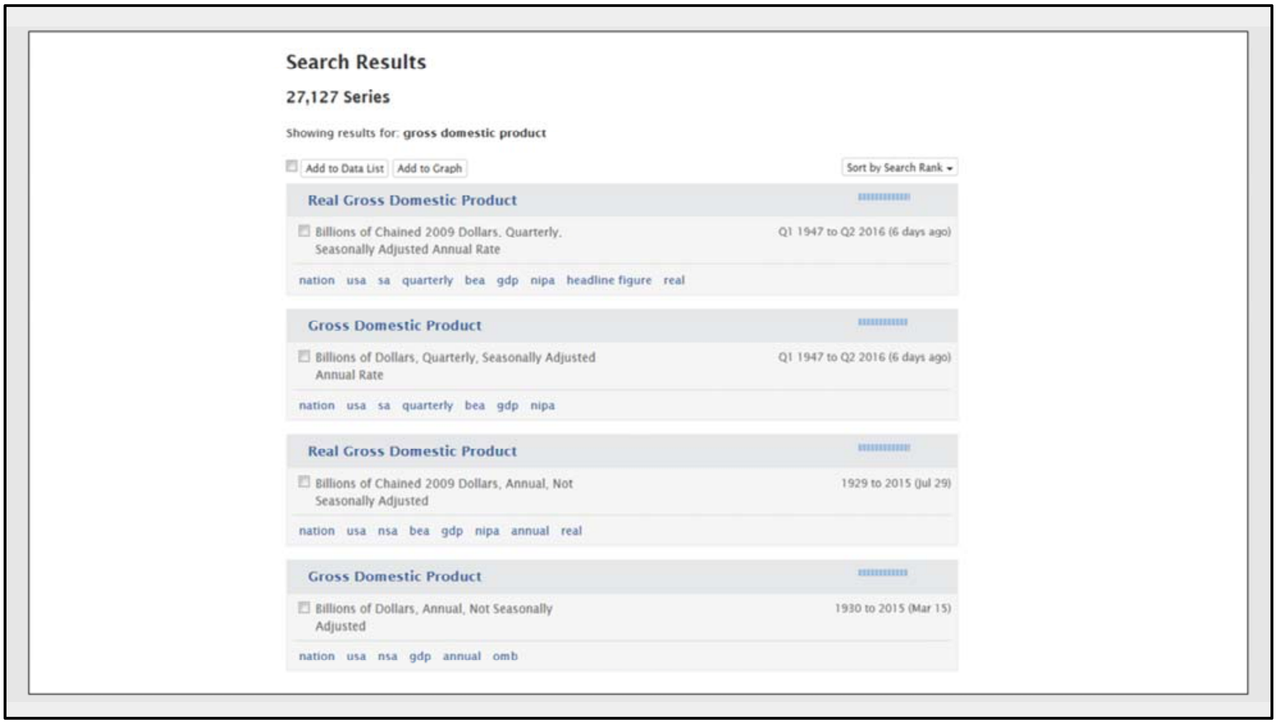
- FRED-S doesn’t have any functionality for capture or transfer of individual data points. FRED and other data providers already have established mechanisms for that function. This schema is exclusively descriptive.



The core problems

- How can I be sure my search results are what I'm really looking for?
- What are the differences between two (or more) series, and what are the similarities?
- How do I know I'm making the correct assumptions about this series?
- How do I avoid easy or common mistakes?
- Where should I look first for help understanding this series?

As I mentioned earlier, FRED's audience is growing – that means more and more people using economic data who don't already have a strong background in economic concepts or vocabulary. There are a lot of easy pitfalls – I've fallen into a lot of them myself – and we want to help users avoid them.



Search Results

27,127 Series

Showing results for: **gross domestic product**

Real Gross Domestic Product

☐ Billions of Chained 2009 Dollars, Quarterly, Seasonally Adjusted Annual Rate — Q1 1947 to Q2 2016 (6 days ago)

nation usa sa quarterly bea gdp nipa headline figure real

Gross Domestic Product

☐ Billions of Dollars, Quarterly, Seasonally Adjusted Annual Rate — Q1 1947 to Q2 2016 (6 days ago)

nation usa sa quarterly bea gdp nipa

Real Gross Domestic Product

☐ Billions of Chained 2009 Dollars, Annual, Not Seasonally Adjusted — 1929 to 2015 (Jul 29)

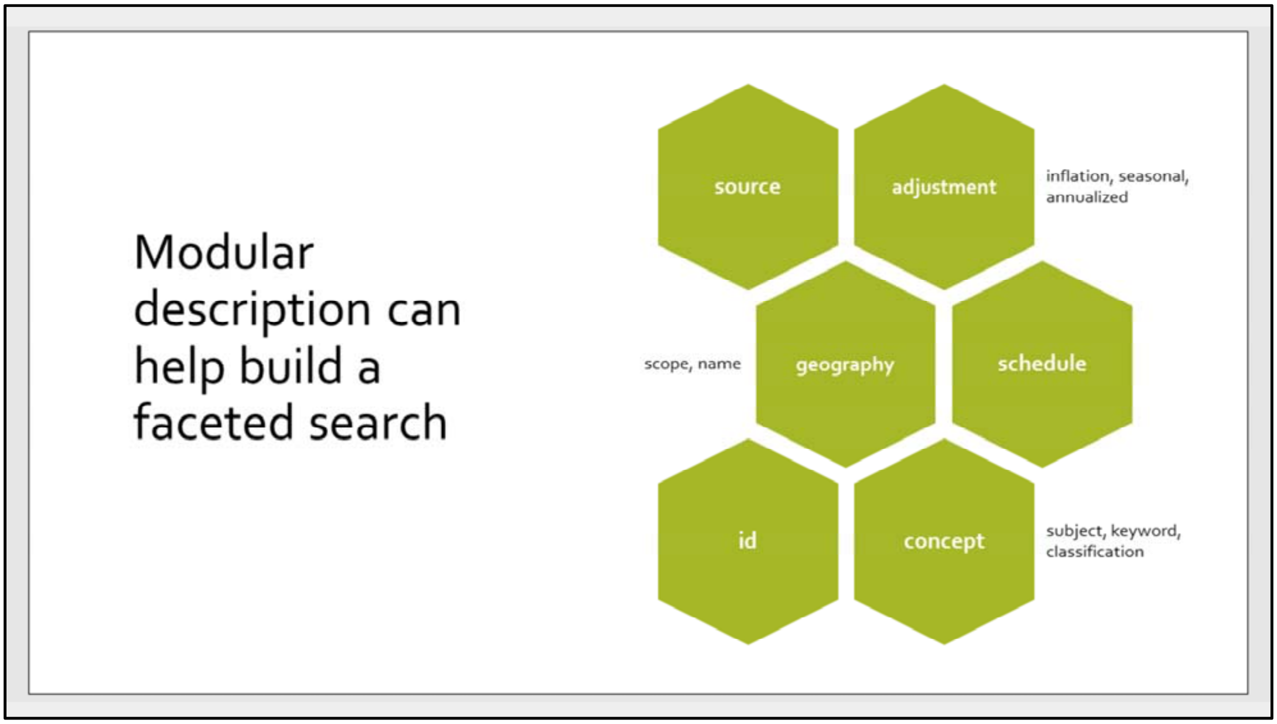
nation usa nsa bea gdp nipa annual real

Gross Domestic Product

☐ Billions of Dollars, Annual, Not Seasonally Adjusted — 1930 to 2015 (Mar 15)

nation usa nsa gdp annual omb

Right now, when you search gross domestic product in FRED, you get more than twenty-seven THOUSAND results. If you don’t already have a strong understanding of how and when and why and by whom the data is put out – let’s say you’re a high school student or a retiree, or, as a COMPLETELY RANDOM example, a librarian with no background in economics – you may be at a loss to distinguish between these series, and pick whatever is on the top of the list.



Modular description can help build a faceted search

source

adjustment

inflation, seasonal, annualized

scope, name

geography

schedule

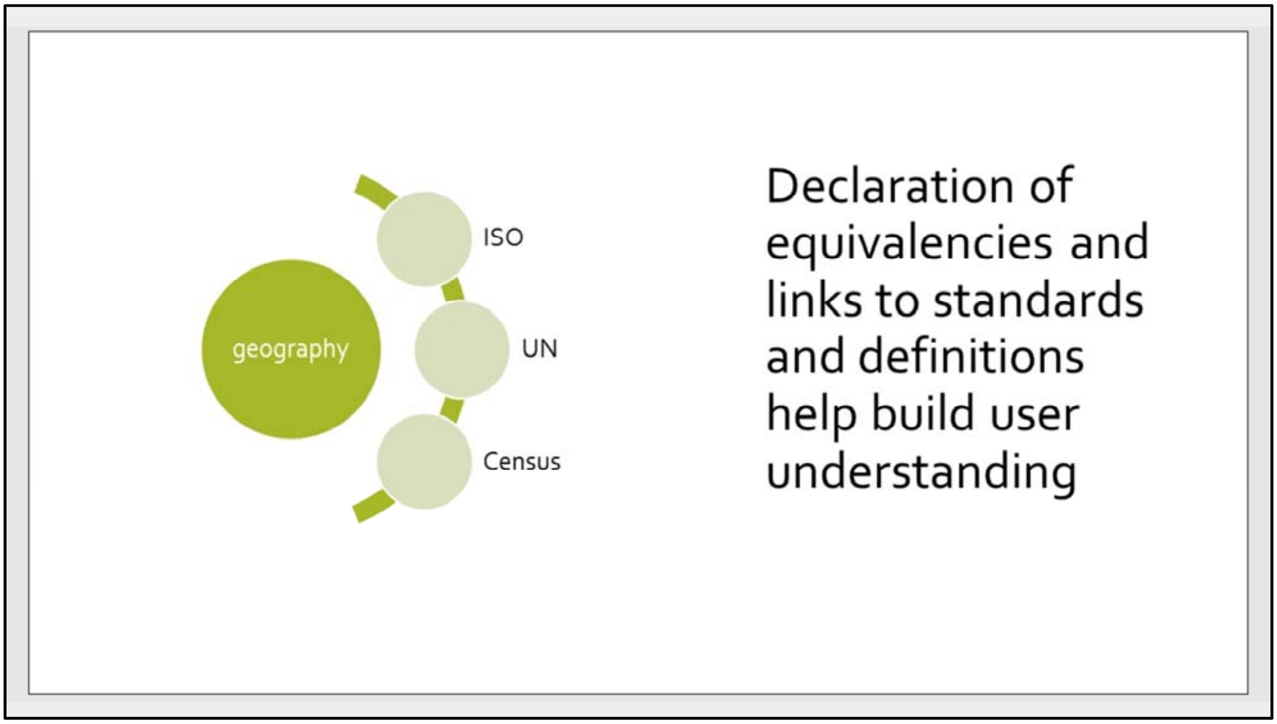
id

concept

subject, keyword, classification

How can I be sure my search results are what I’m really looking for?

FREDS tries to solve this by breaking description into many pieces that can be reassembled into a series title, unit, or identifiers. With modularity, we can build a more sophisticated faceted search



geography

ISO

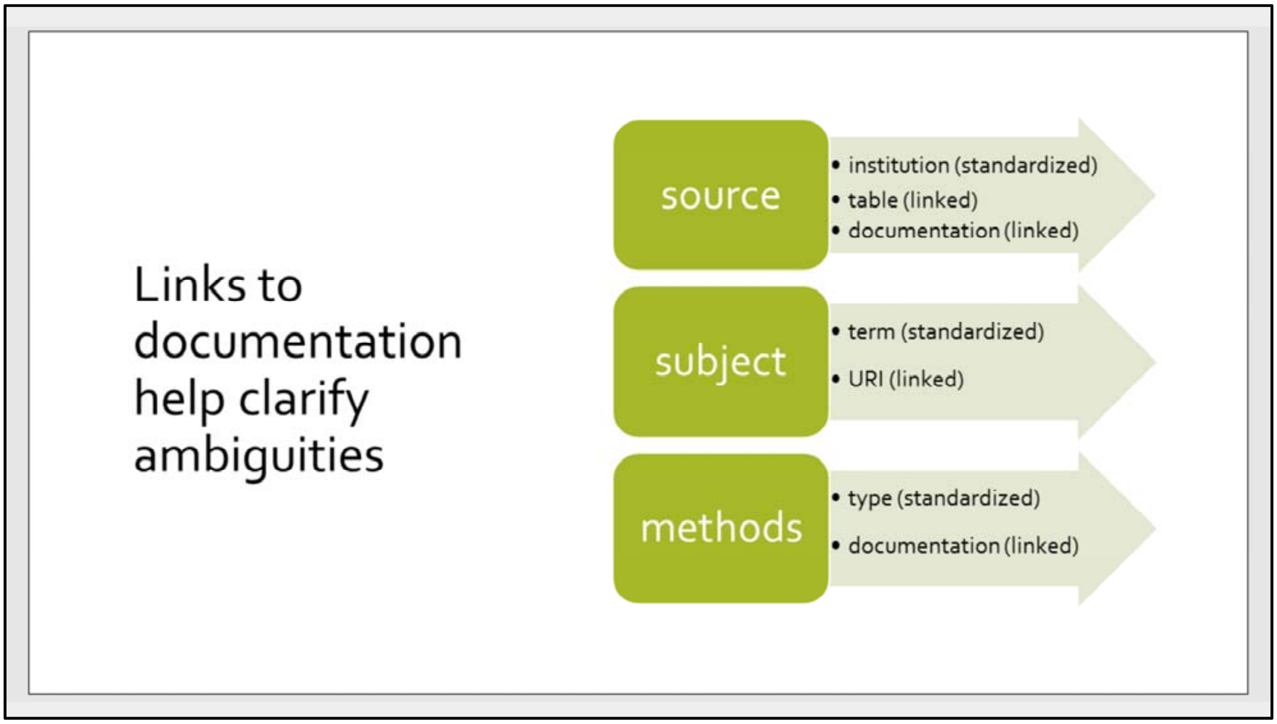
UN

Census

Declaration of equivalencies and links to standards and definitions help build user understanding

What are the differences between two (or more) series, and what are the similarities?

By making use of linked data resources, FRED-S can bring related concepts together. By declaring authorities and standard terms, we can help users figure out if they’re looking at data for the USA, the USA, the USA, or the USA – FRED’s big selling point is the ability to compare and chart multiple series on the same graph, so knowing if you’re looking at data for the 50 states, 50 states plus outlying territories, 48 contiguous states, or all US citizens regardless of location is really helpful!



Links to documentation help clarify ambiguities

source

- institution (standardized)
- table (linked)
- documentation (linked)

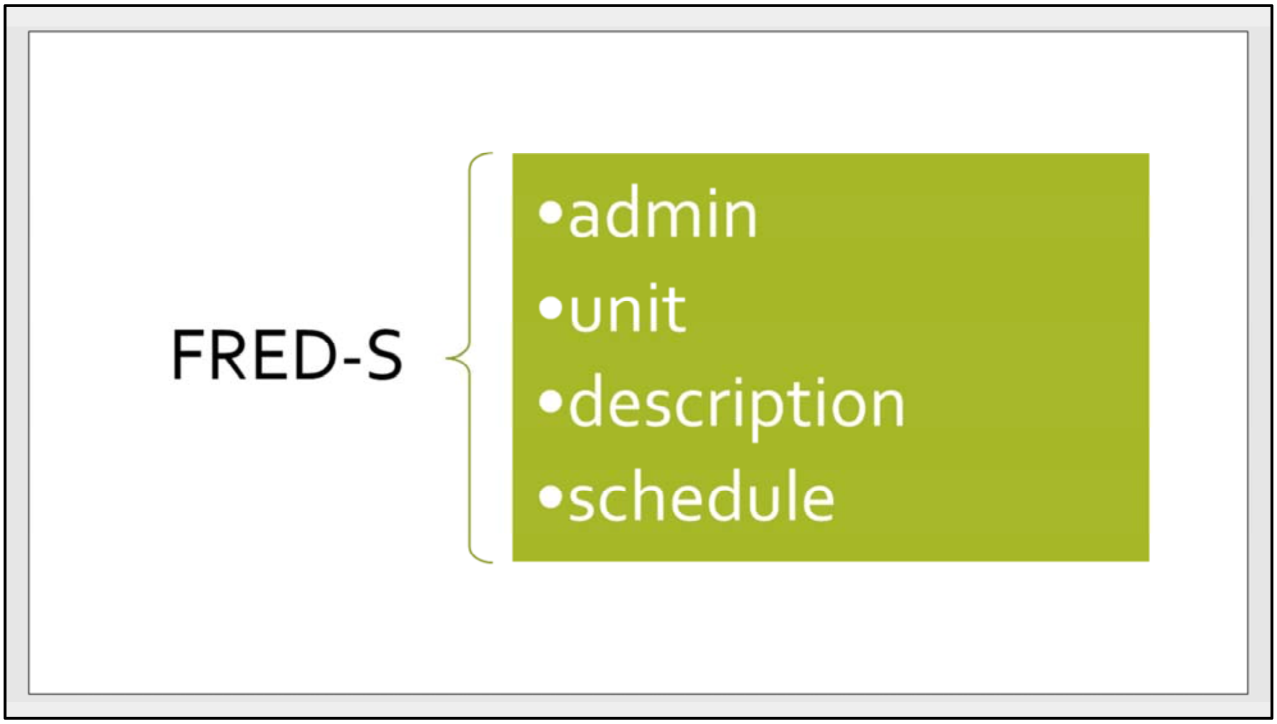
subject

- term (standardized)
- URI (linked)

methods

- type (standardized)
- documentation (linked)

How do I know I’m making the correct assumptions about this series?
How do I avoid easy or common mistakes?
Where should I look first for help understanding this series?

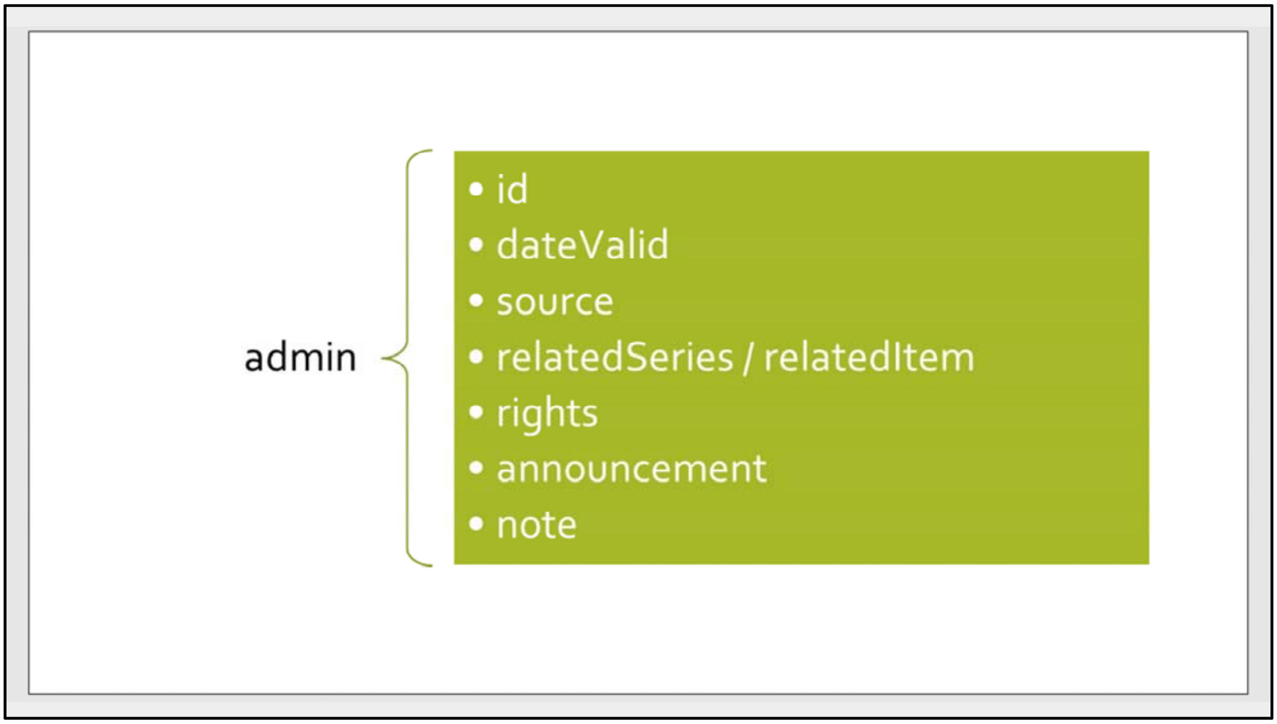


FRED-S

- admin
- unit
- description
- schedule

Briefly, I'm going to walk through the elements that we've identified as central to the description of a time series in FRED and how they've been broken out in the FRED Schema

As I mentioned earlier, there are four core sections or divisions of metadata in FRED-S: admin, unit, description, and schedule

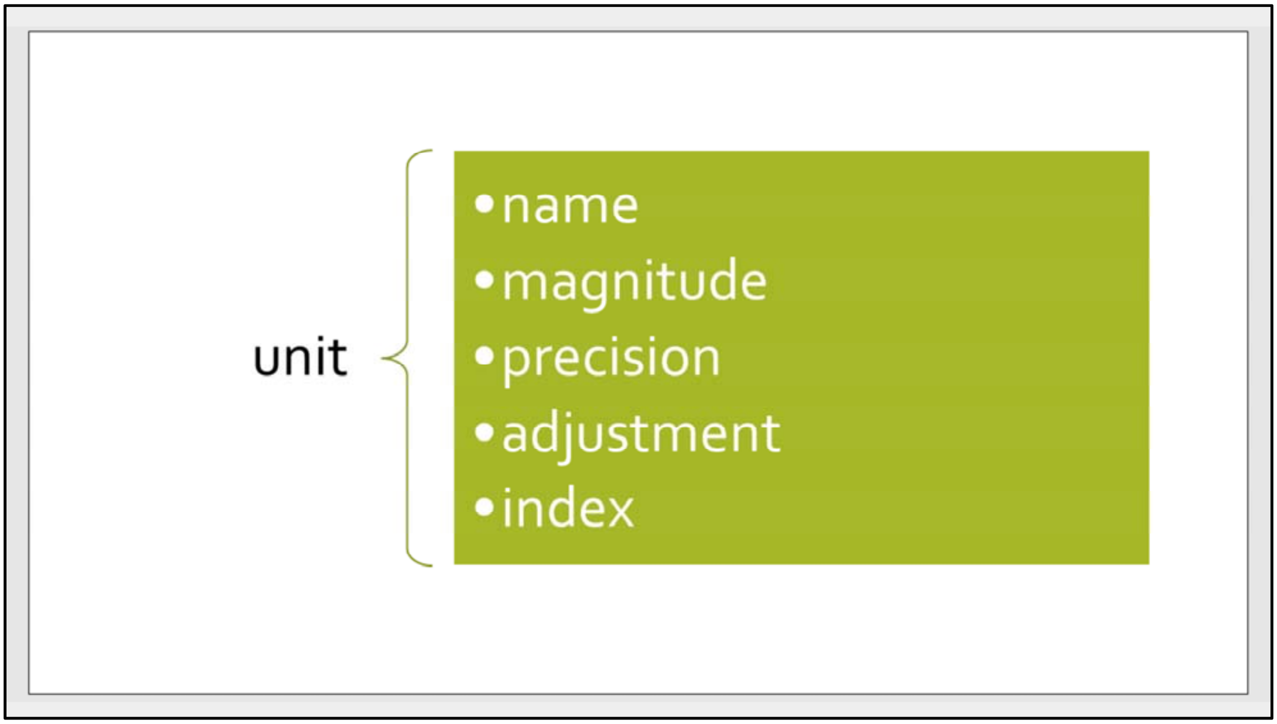


admin

- id
- dateValid
- source
- relatedSeries / relatedItem
- rights
- announcement
- note

There are seven top-level child elements in the "admin" bucket – a series identifier, a valid date for the metadata declaration, the source information (which has additional child elements for source identifier code, institution name, table and release or database), related FRED series and external links, a rights declaration, announcements, and notes.

The admin fields attempt to answer "where does this come from and what's it related to?"

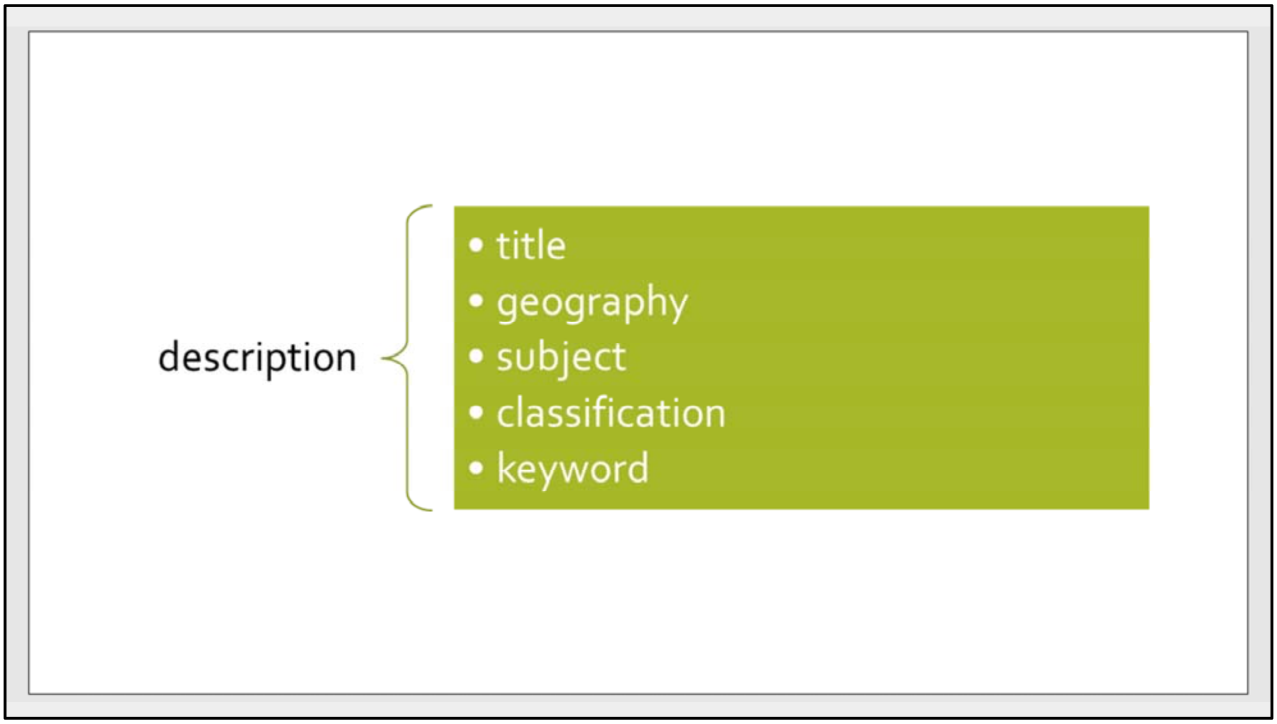


unit

- name
- magnitude
- precision
- adjustment
- index

There are five top-level child elements in the “unit” bucket – the unit tries to break the units down into the core pieces, to allow for more sophisticated transformations, to make standardization easier, and to link to documentation or clarification where needed, as in index

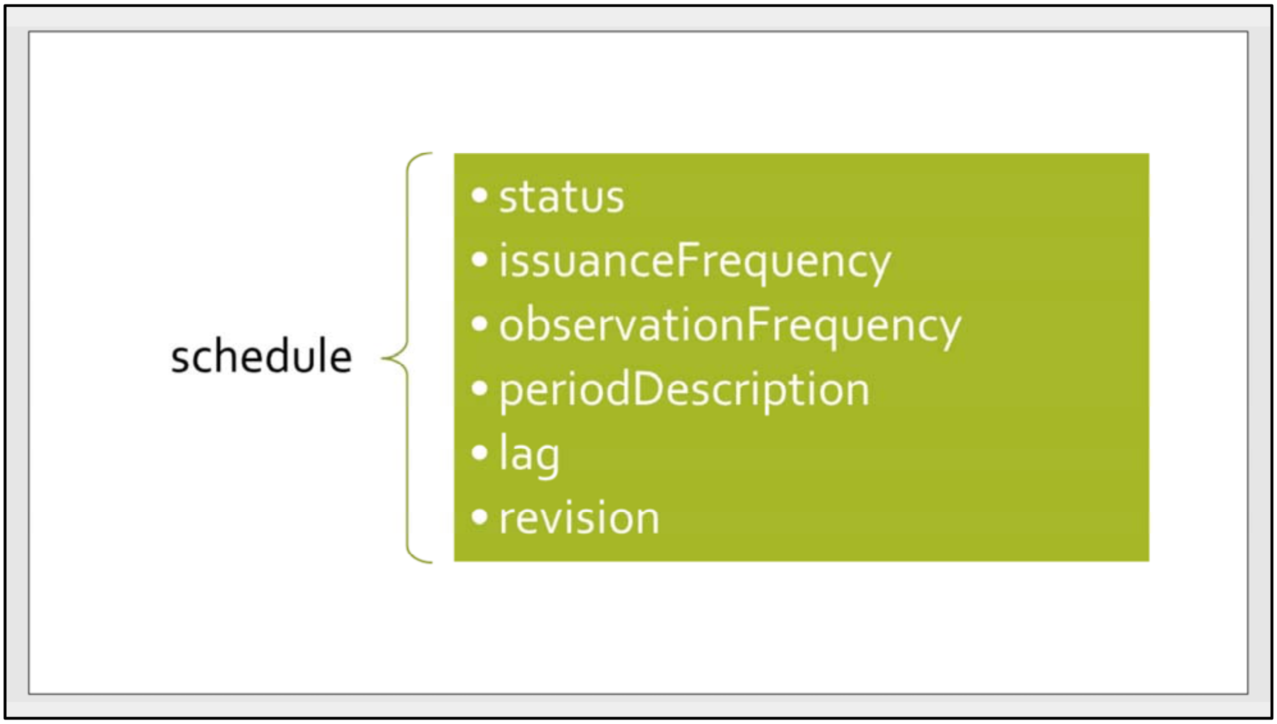
The unit section tries to define the “what” of the measurement - is the series measuring people, dollars, items, an arbitrary value?



description

- title
- geography
- subject
- classification
- keyword

There are five top-level child elements in the “description” bucket, covering the “aboutness” of the series. Geography, subject, and classification link out to authority records, while title and keyword follow an internal style guide.

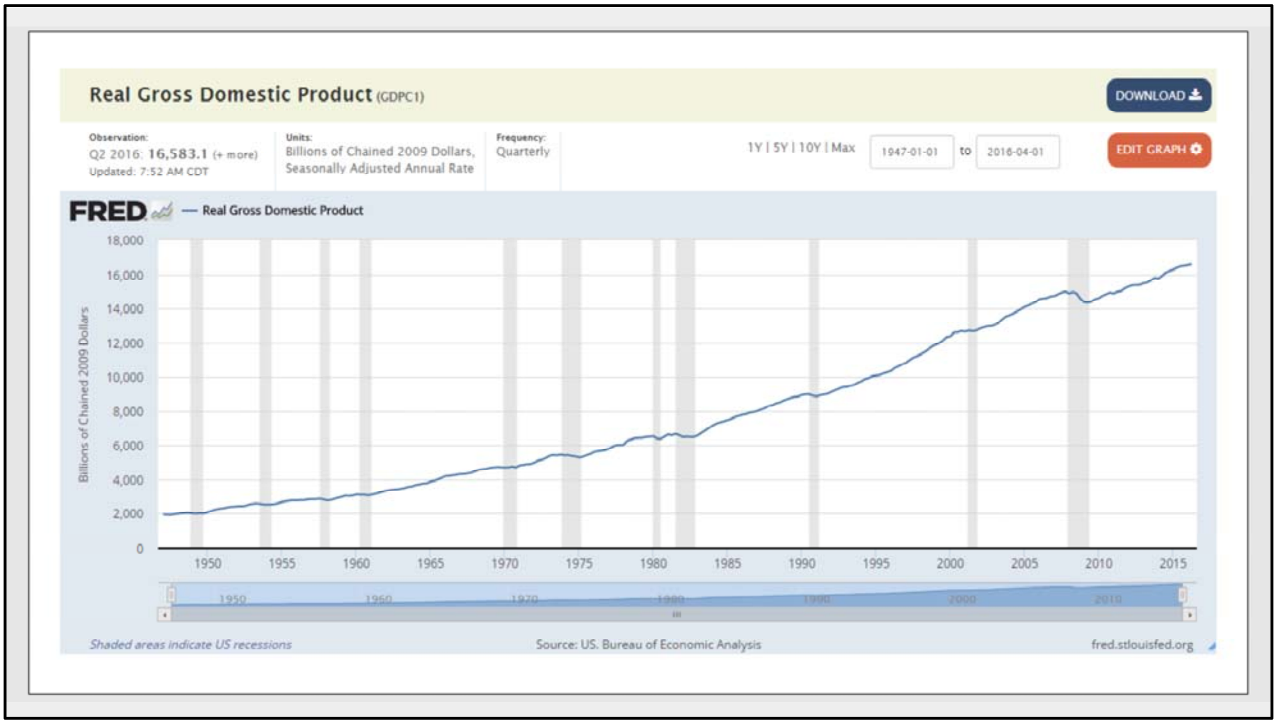


schedule

- status
- issuanceFrequency
- observationFrequency
- periodDescription
- lag
- revision

There are six top-level child elements in the “schedule” bucket, covering publication and update of the series and its data.

As elsewhere in the schema, most of these are designed to both have text values and link back to the documentation or to authority records.



Real Gross Domestic Product (GDPC1)

Observation:
Q2 2016: 16,583.1 (+ more)
Updated: 7:52 AM CDT

Units:
Billions of Chained 2009 Dollars,
Seasonally Adjusted Annual Rate

Frequency:
Quarterly

1Y | 5Y | 10Y | Max 1947-01-01 to 2016-04-01

Shaded areas indicate US recessions

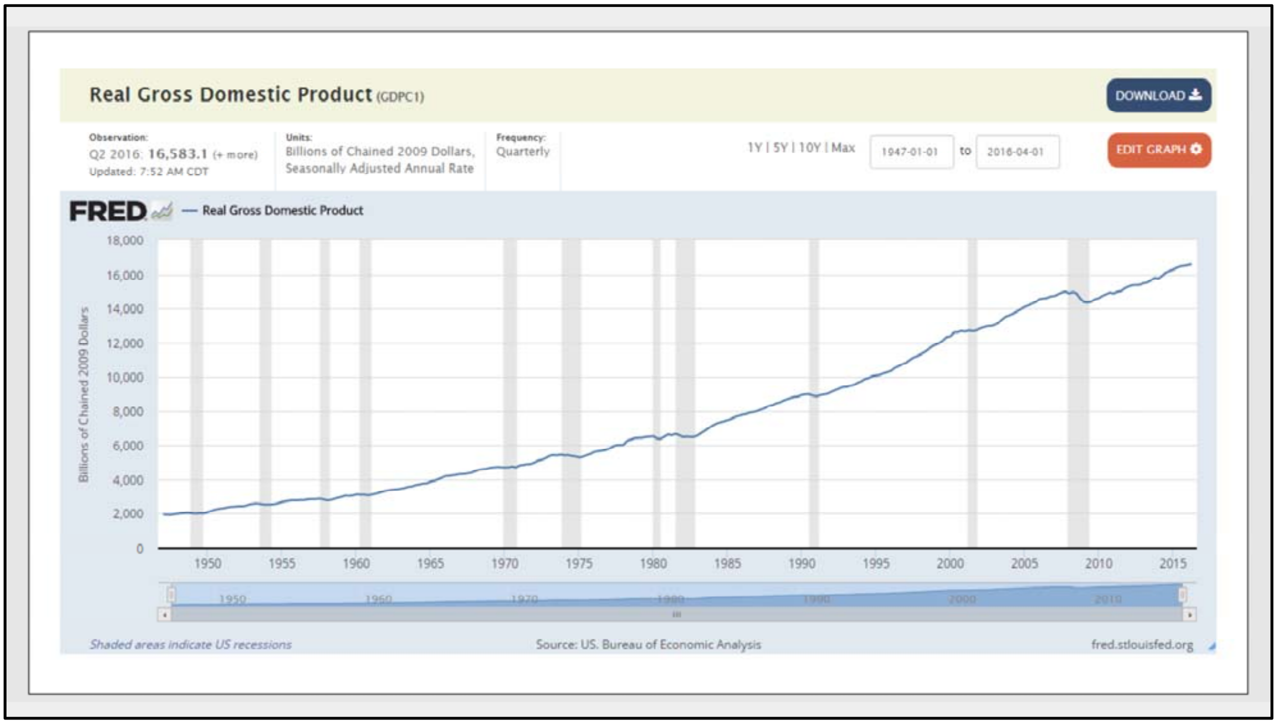
Source: US. Bureau of Economic Analysis

fred.stlouisfed.org

Briefly, here's an example of some common ambiguities and how FRED-S is designed to address them.
Admin – in the title line there, you see the FRED series ID, but there's no built-in metadata for the series code assigned by the source. Historically, the FRED team has gotten around that by making use of free-text note fields, and I'll talk about this a little more in a minute.

Units: as the source does, FRED lists the units as "Billions of Chained 2009 Dollars, Seasonally Adjusted Annual Rate" - on the Y-axis, it's just "Billions of Chained 2009 dollars"

Frequency: FRED lists the observation frequency, but not the release or revision frequency! Unless you already know how GDP is produced and released, you might not be aware that the July, August, and September versions of Quarter 2 2016 are all different – and if you're doing research in a field where historical data is relevant, you might draw some very skewed conclusions.



Real Gross Domestic Product (GDPC1)

Observation:
Q2 2016: 16,583.1 (+ more)
Updated: 7:52 AM CDT

Units:
Billions of Chained 2009 Dollars, Seasonally Adjusted Annual Rate

Frequency:
Quarterly

1947-01-01 to 2016-04-01

Shaded areas indicate US recessions

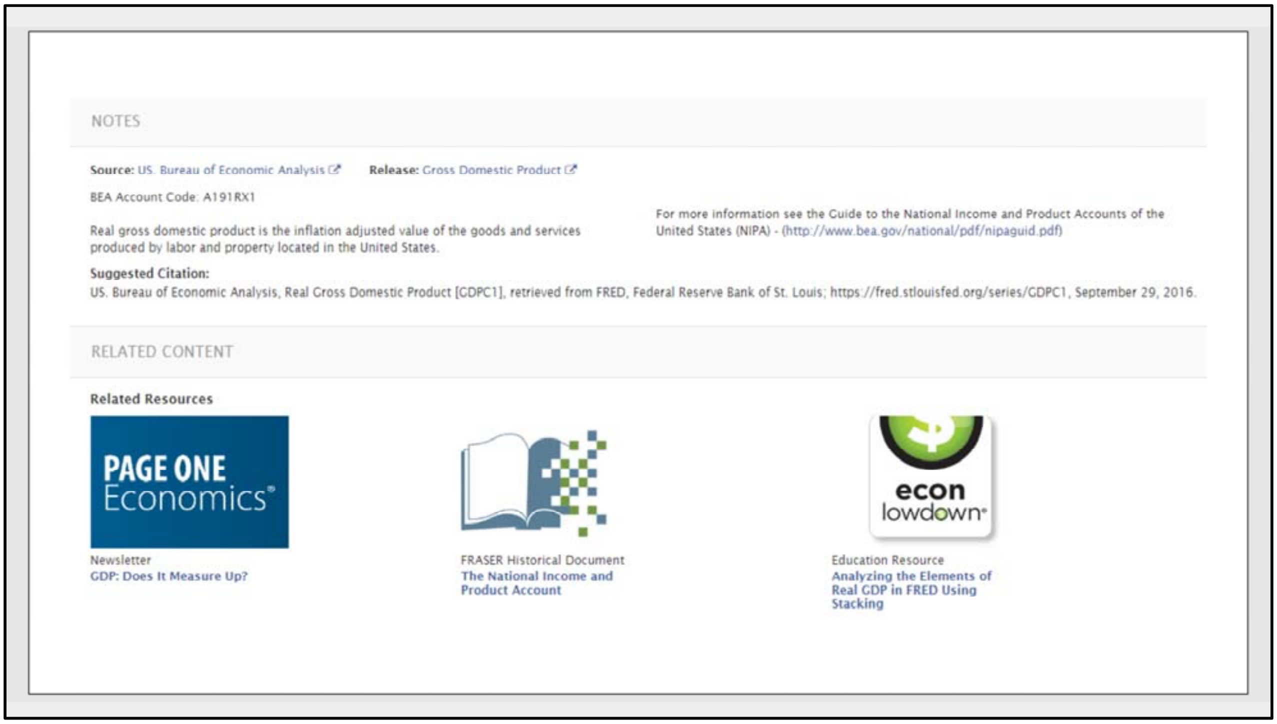
Source: US. Bureau of Economic Analysis

fred.stlouisfed.org

Let’s talk briefly about the units.

As I mentioned, FRED lists the units as “Billions of Chained 2009 Dollars, Seasonally Adjusted Annual Rate” in the top metadata, but on the Y-axis, it’s just “Billions of Chained 2009 dollars”. If you know the source tables for this data you know that those are both correct, depending on how you read the table or what you consider a “unit” to be.

Using FRED-S, we can break that into a core unit of “U.S. Dollars”, then declare an inflation adjustment with an identified base year and a link to the methodology, declare a seasonality adjustment with a link to methodology, and declare an annualization with a link to the methodology. None of that **has** to be displayed on the graph itself, but we can make that metadata available to the public, and in the future we can build an interface that can accommodate that information.



NOTES

Source: US. Bureau of Economic Analysis **Release:** Gross Domestic Product

BEA Account Code: A191RX1

Real gross domestic product is the inflation adjusted value of the goods and services produced by labor and property located in the United States.

For more information see the Guide to the National Income and Product Accounts of the United States (NIPA) - (http://www.bea.gov/national/pdf/nipaguid.pdf)

Suggested Citation:
US. Bureau of Economic Analysis, Real Gross Domestic Product [GDPC1], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/GDPC1, September 29, 2016.

RELATED CONTENT

Related Resources

Newsletter
GDP: Does It Measure Up?

FRASER Historical Document
The National Income and Product Account

Education Resource
Analyzing the Elements of Real GDP in FRED Using Stacking

So here’s where our administrative metadata, description, and related content has been presented – Source name with a linkout, release name with a linkout, and any other metadata is handled by notes – the source account code, as I mentioned earlier, is a simple text note. On the FRED back-end, all notes are currently actually included in a single text field with line breaks. With FRED-S, we can enrich that information.

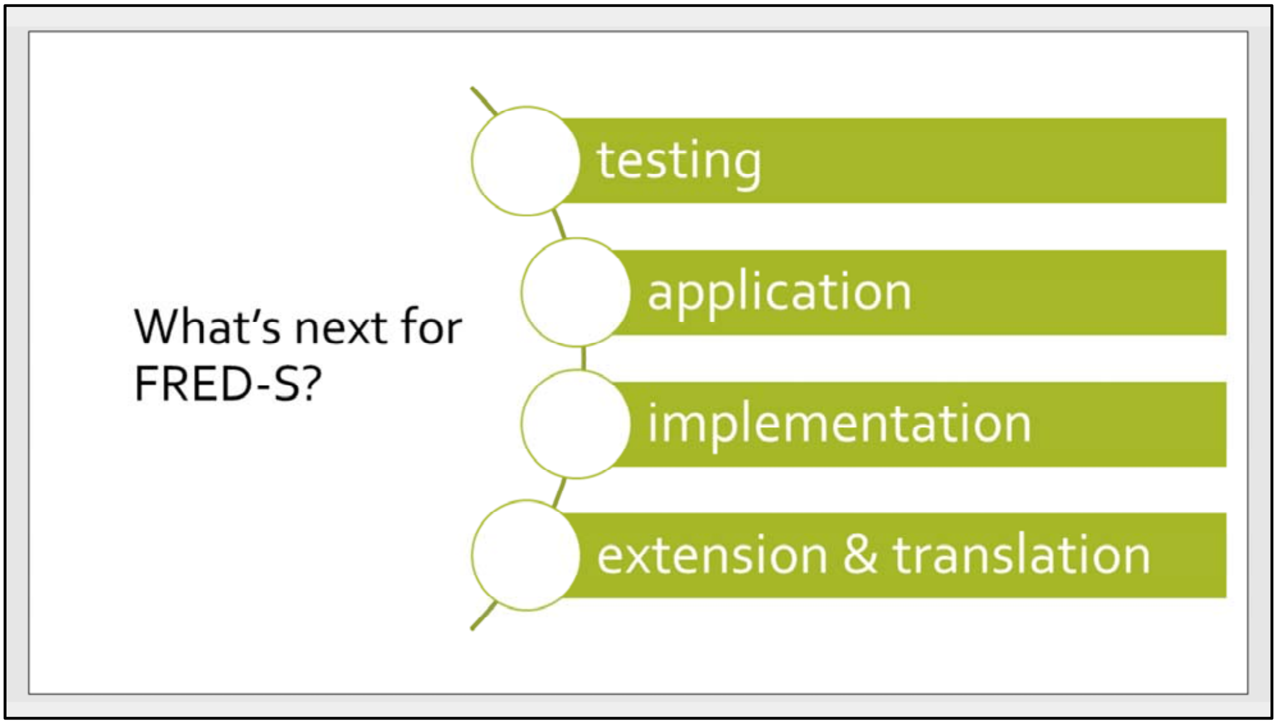
These related resources are manually encoded now, as well. The admin fields in FRED-S will allow us space to clearly and explicitly link to specific resources like these, but also hopefully automatically generate related links.

As far as description goes, FRED has traditionally not listed a geography for any U.S. data from a U.S. source – but I can personally attest to the fact that it’s easy to miss the fact that the BEA doesn’t use the ISO definition of the USA, because the ISO definition includes outlying territories. Until I was writing the geography documentation for the schema, it didn’t even occur to me that national data might define the nation differently.

One more thing we think that’s really important to encode that hasn’t been addressed in FRED in the past is the declaration of rights metadata. There’s often confusion about where this data comes from and to whom it belongs – as Keith Taylor mentioned during the data producers panel on Thursday, people occasionally attribute everything in FRED to the St.

Louis Fed – presentation is conflated with ownership. And remember, the regional Federal Reserves aren't government entities; although we provide access to a lot of data that's public domain, our actual output isn't automatically public domain. By requiring a rights statement for every series – even if it's "copyright not evaluated" – we aim to help clear up that particular ambiguity.

NEXT STEPS



What’s next for FRED-S?

- testing
- application
- implementation
- extension & translation

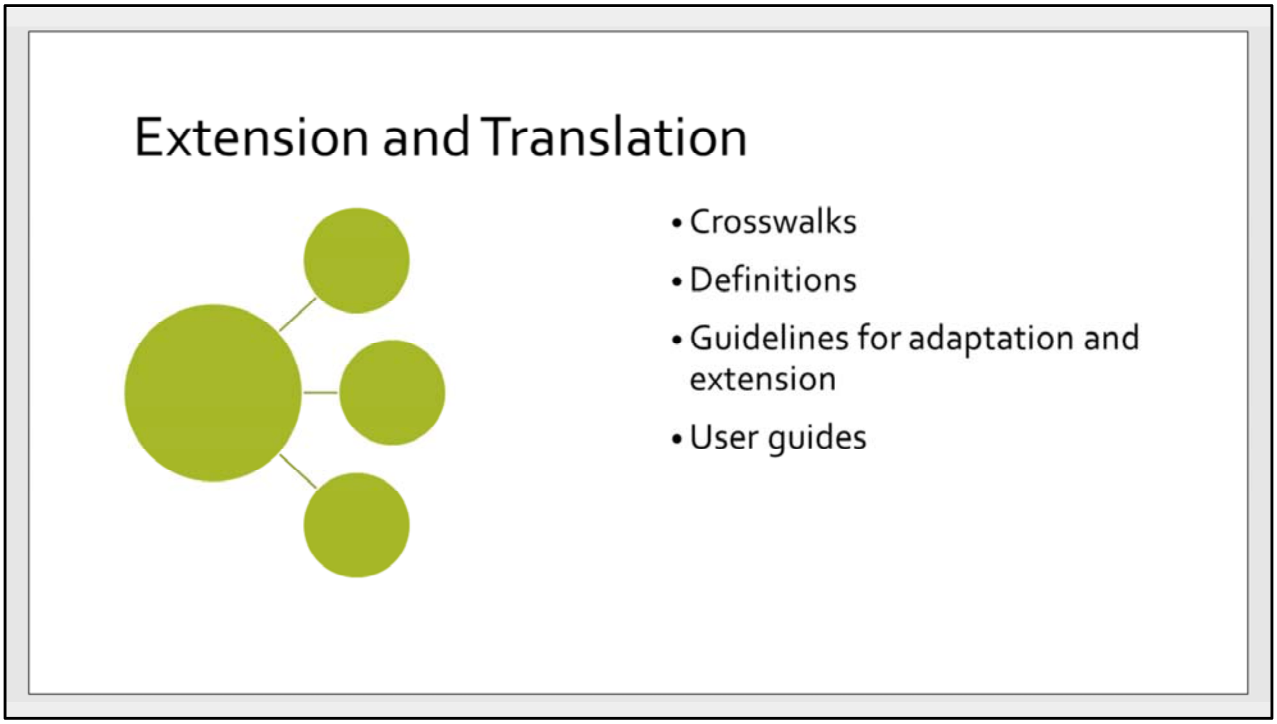
So what’s next?
First is testing – we’ve asked data librarians within the Federal Reserve system to help us fine-tune the schema to find where it does and does not meet their patron’s needs. We’re also reaching out to data producers to see where we can reuse their metadata for FRED-S in an automate-able way

The other three will likely be in process simultaneously:

Application – documenting FRED series using the schema. We hope that this will expose pain points for data entry tasks and find additional areas for eventual extension of the schema

Implementation – working with our web developers, and with focus groups of internal users, we’re going to identify and start to build ways where this metadata can be used and displayed to enhance the user experience;

And Extension and translation – Building a release-level FRED-S guidelines, building crosswalks to DDI and SDMX, and identifying other areas where additional elements or documentation might be needed to effectively capture all the relevant information about these series



Extension and Translation

- Crosswalks
- Definitions
- Guidelines for adaptation and extension
- User guides

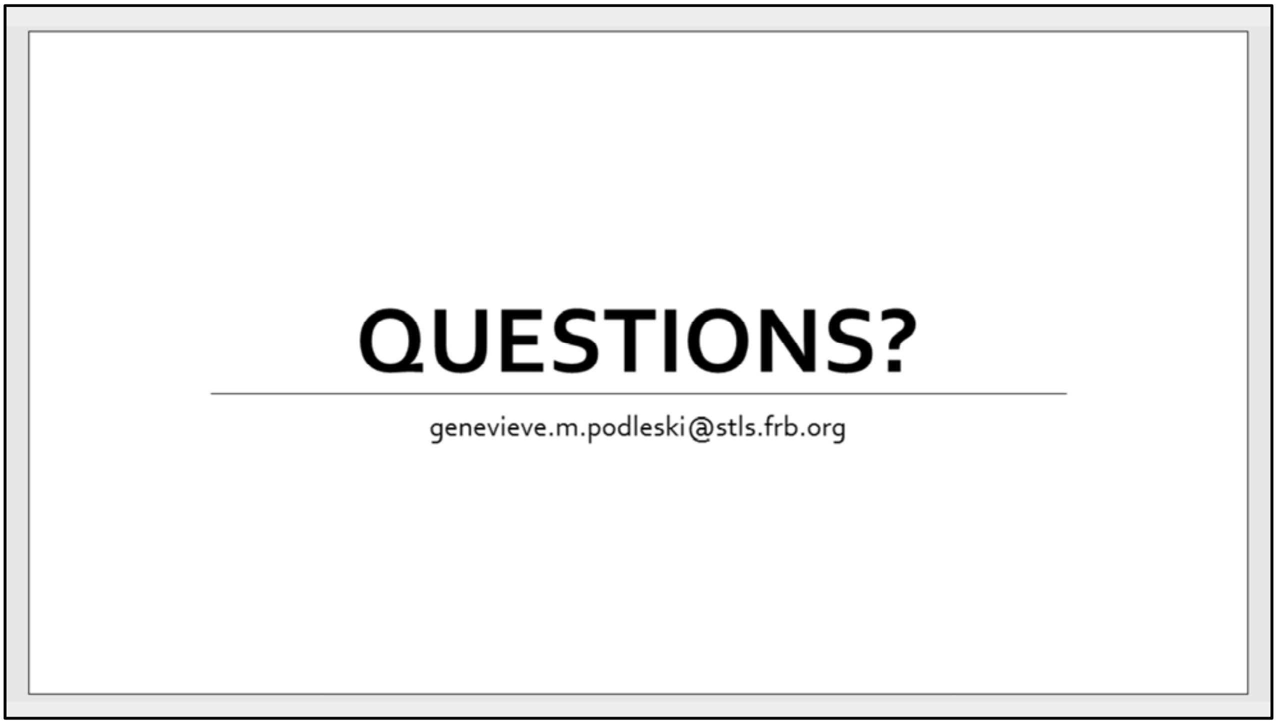
Just to elaborate on this extension and translation part –

We know what while we're working on trying to solve this problem for FRED, our counterparts at data producing institutions and other aggregators, commercial or nonprofit, as well as academic and data librarians, are working on their own initiatives and adapting preexisting standards for their own work.

Our intention has always been to make this schema flexible and extensible to other use cases in time series data. We have already "extended" what we consider the core schema into a few FRED-specific elements and attributes, like a specific field for documenting how much of an embargo we have between the release of a series and when it's posted on FRED.

At no point do we intend to pick one authoritative source or one set of terms – if it's documented (or documentable), we want to use it and make it more usable.

We hope that other folks in the metadata community will find this useful and help us identify new ways to use and adapt FRED-S as we move forward.



QUESTIONS?

genevieve.m.podleski@stls.frb.org