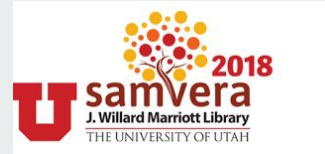# Samvera
The Change we believed in

Developing the United States Department of Agriculture's Economics Statistics and Market Information System (USDA ESMIS) In Hyrax
https://usda.library.cornell.edu

Tahir Poduska, Lynette Rayle
Cornell University

Samvera Connect 2018
Salt Lake City, UT

About USDA ESMIS

- Partnership between Mann Library at Cornell University and five agencies of the U.S. Department of Agriculture (USDA).
- Repository for over 2100 publications and thousands of associated releases covering U.S. and international agriculture and related topics.
- Subscription and email notification service for latest releases of active publications
- Around 60,000 subscribers in the existing system
- Close to a million email send per month for new releases

Cornell University

USDA ESMIS Project Team

Ochs, Mary
Director of Mann Library

Poduska, Tahir
Web Developer

Fereira, John
Technology Strategist

Shea, Ashley
Food & Agriculture Librarian &
Instruction Coordinator

Delisle, Greg
DevOps Engineer

Branchini, Darcy
Manager of Land Grant IT

Kidwell, Kevin
User Experience Developer

Rayle, Lynette
Application Developer

Bryant, Selena
User Engagement & Inclusion
Librarian

Manly, Chris
Director of Library Systems

samvera

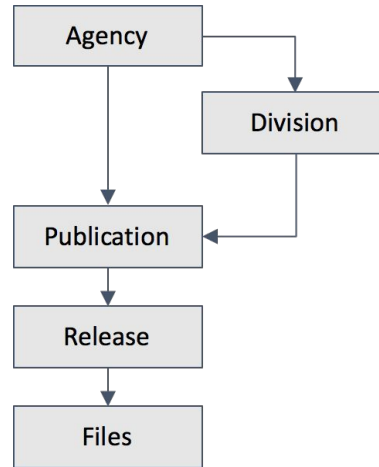# PCDM Modelling of USDA ESMIS in Hyrax

Agency-> Collection

Division-> Nested Collection
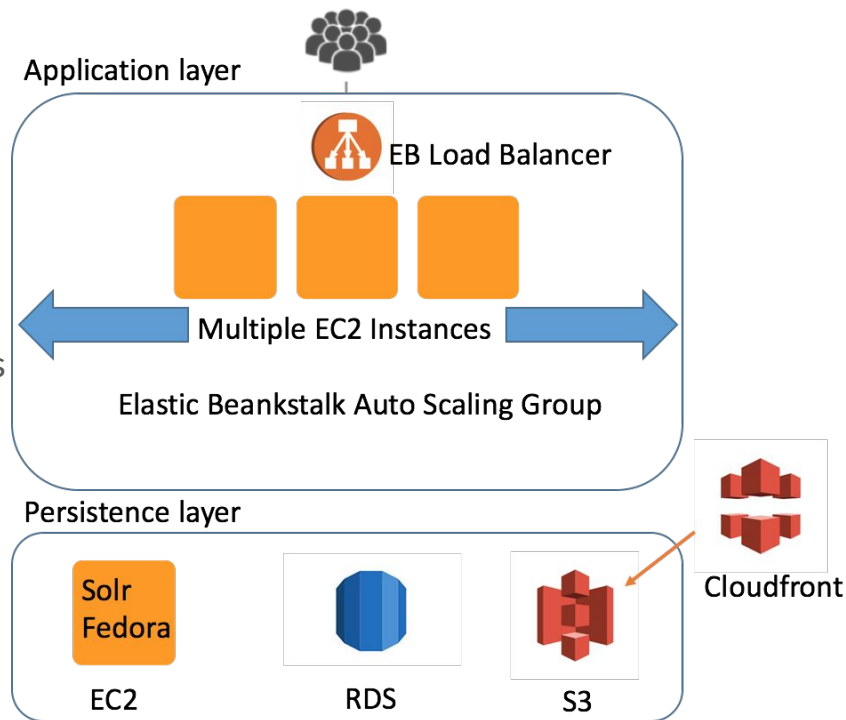
Publications-> Work

Releases -> Child work of Publication
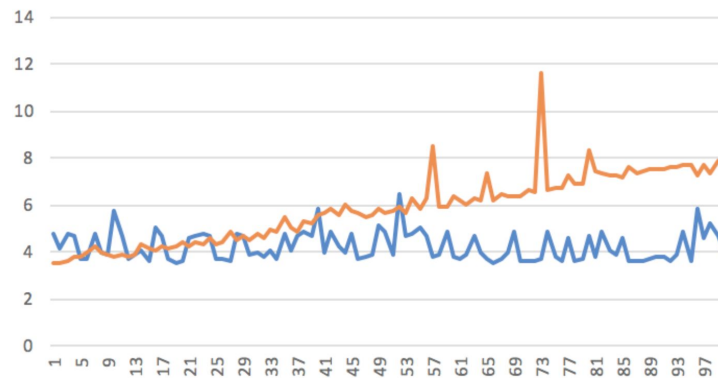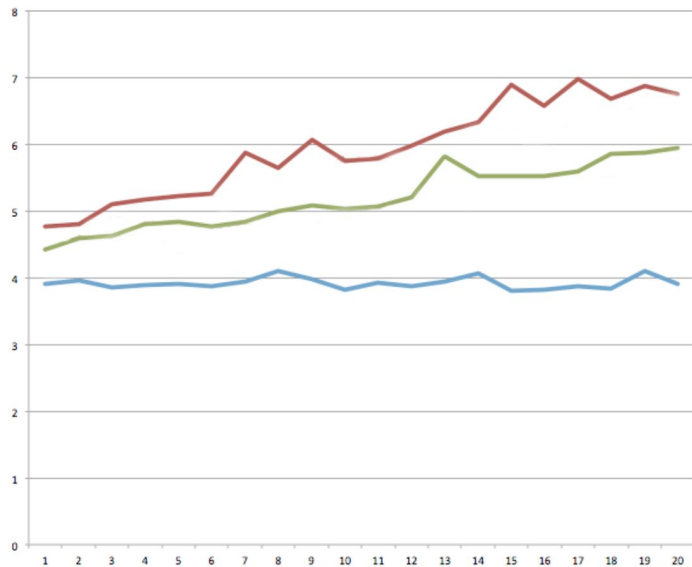
Release Binaries -> FileSets

Infrastructure & Workflow

- AWS Elastic Beanstalk for Rails App (Hyrax)
- EC2 instance for Fedora and Solr
- S3 for file storage with Web hosting
- and CloudFront for file access
- Shoryuken with SQS queues for background jobs
- GiHub PRs, Jenkins Integration for CI
- Monitoring and Logging: AppSignal, PaperTrail

Application layer

EB Load Balancer

Multiple EC2 Instances

Elastic Beankstalk Auto Scaling Group

Persistence layer

Cloudfront

Solr
Fedora

EC2

RDS

S3

Graphs

Customization

- Agencies and Divisions: From AdminSets to Collection Extensions
- User Informed Design: Custom date range filters, Forms
- File Storage and Access: AWS S3 and CloudFront
- USDA ESMIS API: Adapt Hyrax WorksControllerBehavior
- Subscription and Email: MailGun API
- Performance Improvements

Customizations for Agencies and Divisions

- Our design required Agencies to be able to have Divisions nested under them

- Started work under Hyrax 2.0 which does not allow nesting of collections

Take One: Agencies and Division as Admin Sets

- Make all agencies and divisions admin sets

- Add fields to admin sets
  - Unit type: AGENCY or DIVISION
  - Parent: id of parent Agency or nil

Seems simple enough on the surface, but not really

Take One: Agencies and Division as Admin Sets

Challenges: Many many customizations that are fragile.

- Have to customize all occurrences of the Admin Set label to be Agency or Division
- Have to customize the Admin Set edit form to have the Agency metadata or the Division metadata
- Have to customize Admin Set search builders to…
  - return just agencies
  - return just divisions associated with an agency
- And many others

Became evident that these customizations were becoming more and more ugly.

Time for Plan B

Take Two: Agencies and Divisions as Collections

- Very briefly considered trying to make similar customizations to Collections

- The energy around Collection Extensions was building which would include several

  enhancements we were interested in

  - nesting
  - sharing - allowing teams to deposit publications in an agency or division
  - applying permissions from collection to work

- We decided to dedicate our time to moving forward the Collection Extensions work

Much Time Passes

Take Three: Agencies and Divisions as Collections

- Super easy to create and nest Agencies and Divisions

- Super easy to set up groups for each Agency and each Division of only the users who can edit and deposit for each

- Super easy to have those same groups be the editors and viewers of the works created in those Agencies and Divisions

Still had to do some customizations but they are simpler, more inline with Hyrax normal operations, and will be easier to maintain over time.

Success

Customization: Date Range Facets

Browse by:

Category >

Release Date ⌄

Last Month   28
📅 Custom Range   451

‹	Oct 2018										Nov 2018		›
Su	Mo	Tu	We	Th	Fr	Sa		Su	Mo	Tu	We	Th	Fr	Sa
30	1	2	3	4	5	6		28	29	30	31	1	2	3
7	8	9	**10**	11	12	13		4	5	6	7	8	9	10
14	15	16	17	18	19	20		11	12	13	14	15	16	17
21	22	23	24	25	26	27		18	19	20	21	22	23	24
28	29	30	31	1	2	3		25	26	27	28	29	30	1
4	5	6	7	8	9	10		2	3	4	5	6	7	8

« Previous | **1 - 10** of **2,113** | Next »

Sort by relevance ⌄   10 per page ⌄

NASS Livestock Slaughter

Log in or Sign Up to subscribe

DESCRIPTION:

This monthly full-text file contains the number of head, live weight, dressed weight of cattle, calves, sheep, lambs, hogs and pigs slaughtered in commerical plants and number by class in federally inspected plants by regions as well as red meat production by species by states and U.S.. Data for this publication comes from reports completed by the Food Safety and Inspection Service (FSIS), and USDA, combined with data from State-administered Non-Federally Inspected (NFI) slaughter plants.

Animals and Animal Products:Livestock

pork, beef, production, slaughter, veal, and mutton

Sep 20, 2018 PDF | TXT | ZIP ℹ

Log in or Sign Up to subscribe

This monthly report contains the number of broiler chicks placed and eggs set for the previous week for 19 selected states and the United States. Included alongside the data is some commentary of observed trends and fluctuations in broiler hatchery statistics within the past month.

Animals and Animal Products:Poultry

Customization: Date Range Filters

Broiler Hatchery NASS

This monthly report contains the number of broiler chicks placed and eggs set for the previous week for 19 selected states and the United States. Included alongside the data is some commentary of observed trends and fluctuations in broiler hatchery statistics within the past month.

Frequency of Release: Weekly

Upcoming releases: Nov 14 2018 3:00 PM, Dec 5 2018 3:00 PM, Oct 17 2018 3:00 PM, Nov 21 2018 12:00 PM, Aug 1 2018 3:00 PM, Dec 27 2018 3:00 PM, Aug 8 2018 3:00 PM, Sep 12 2018 3:00 PM, Sep 6 2018 3:00 PM, Sep 19 2018 3:00 PM, Oct 24 2018 3:00 PM, Nov 7 2018 3:00 PM, Dec 19 2018 3:00 PM, Oct 10 2018 3:00 PM, Sep 26 2018 3:00 PM, Aug 15 2018 3:00 PM, Dec 12 2018 3:00 PM, Aug 22 2018 3:00 PM, Aug 29 2018 3:00 PM, Nov 28 2018 3:00 PM, Oct 3 2018 3:00 PM, Oct 31 2018 3:00 PM

Category: Animals and Animal Products:Poultry

Keywords: hatcheries, chicks, eggs, poultry

LATEST RELEASE
Sep 19, 2018 PDF TXT ZIP ⓘ

PUBLICATION CONTACT

Contact office name: National Agricultural Statistics Service
Contact email: nass@nass.usda.gov

Subscribe

Releases

📅 **Filter by Date**

Date	Download			
Sep 19, 2018	PDF	TXT	ZIP	
Sep 12, 2018	PDF	TXT	ZIP	
Sep 6, 2018	PDF	TXT	ZIP	
Aug 29, 2018	PDF	TXT	ZIP	
Aug 22, 2018	PDF	TXT	ZIP	

Today
Yesterday
This Week
Last Week
This Month
Last Month
Custom Range

‹ Oct 2018 ›   ‹ Nov 2018

Su	Mo	Tu	We	Th	Fr	Sa		Su	Mo	Tu	We	Th	Fr	Sa
30	1	2	3	4	5	6		28	29	30	31	1	2	3
7	8	9	10	11	12	13		4	5	6	7	8	9	10
14	15	16	17	18	19	20		11	12	13	14	15	16	17
21	22	23	24	25	26	27		18	19	20	21	22	23	24
28	29	30	31					25	26	27	28	29		

Customization: Date Range Filters

Bootstrap Date Range Picker to filter publications and releases by date

- **Search Results Page:**
  - *class SearchBuilder < Blacklight::SearchBuilder*
    - Publications filtered by release date range using
- **Publication show page:**
  - *class PublicationReleasesSearchBuilder < ::SearchBuilder*
    - Releases Child members filtered by date range

Customization: Multi-Step Forms

- **jQuery SmartWizard**
- **Custom JS Validation**

Customization: File Storage and Access

- **Files in S3, outside Fedora:**
  - *Hyrax::JobIoWrapper*
    - extracted_mime_type
- **File Download from S3:**
  - *Hyrax::DownloadsController*
    - Download file from s3
  - AWS CloudFront serves download requests to s3

NASS Livestock Slaughter

DESCRIPTION: This monthly full-text file contains the number of head, live weigh cattle, calves, sheep, lambs, hogs and pigs slaughtered in comme by class in federally inspected plants by regions as well as red me by states and U.S.. Data for this publication comes from reports co Safety and Inspection Service (FSIS), and USDA, combined with d administered Non-Federally Inspected (NFI) slaughter plants.

CATEGORY: Animals and Animal Products:Livestock

KEYWORD: pork, beef, production, slaughter, veal, and mutton

LATEST RELEASE: Sep 20, 2018 PDF | TXT | ZIP ⓘ

Properties

ebucore: **filename**
release.pdf

ebucore: **hasMimeType**
message/external-body;access-type=URL;url="http://usda-data-dev.s3-website-us-east-1.amazonaws.com/usda-esmis/files/gm/80/hv/32/gm80hv32k/release.pdf"

fedora: **created**
2018-10-10T17:27:55.238Z

Customization: USDA ESMIS API

- **Read API: Endpoints to get metadata for Agencies, Publications, and Releases**
  - ActiveFedora::SolrService.query
    - Filter query by the model, Filter list for the fields to output
    - Clean up all the solr suffixes (ssim, tesim)
- **Write API:** Endpoints to create and updated Publications, and Releases
    - Hyrax::WorksControllerBehavior
    - Assign attributes and relationships for ESMIS (e.g. agency/division as member of collection attributes) from parameters
- **Knock Gem:** JWT authentication for API
- **API Documentation:** Swagger UI
- **Migration**: Used API to migrate >2000 publications and >30,000 releases

Customization: USDA ESMIS API

publication Services about publications		⌄
POST /publication/create Create a Publication		🔒
POST /publication/update Update a Publication		🔒
GET /publication/findAll Returns publications		🔒
GET /publication/findByAgency/{agencyID} find Publication by Identifier		🔒
GET /publication/findById/{id} find Publication by ID		🔒
GET /publication/findByIdentifier/{identifier} find Publication by Identifier		🔒
GET /publication/search Search Publications by Term		🔒
release Services about releases		⌄
POST /release/create Create release		🔒
POST /release/update Update release		🔒
GET /release/findById/{releaseID} find release by releaseID		🔒
GET /release/findByIdentifier/{identifier} find releases by identifier		🔒

Customization: Subscription and Emails
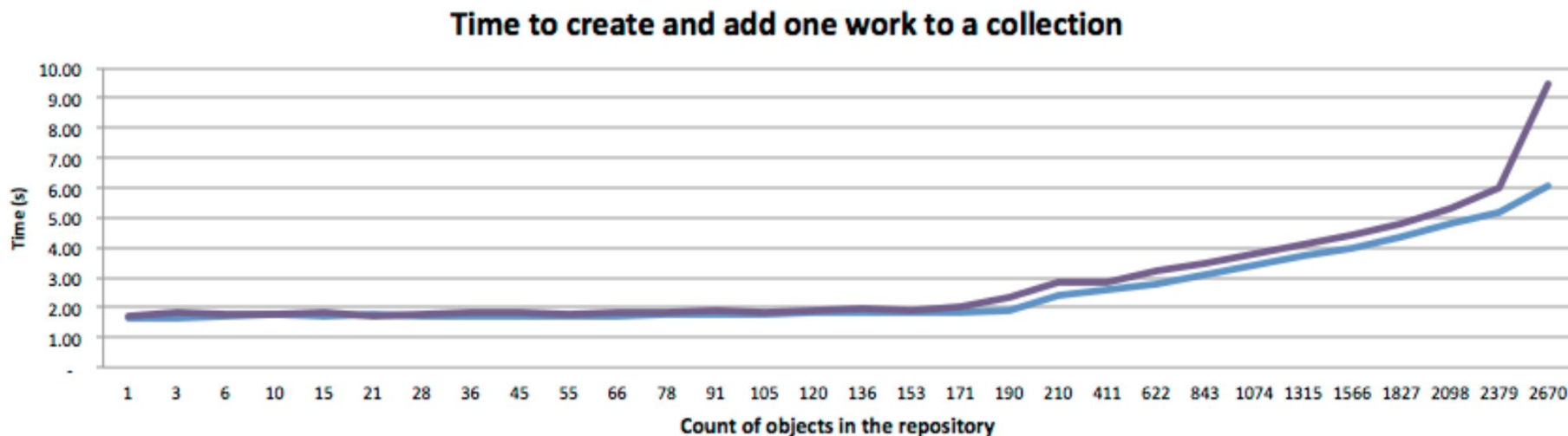
Customization: Subscription and Emails

- **User subscriptions for publications:**
  - *class User*
    - *Added subscriptions attribute (Array of ids) to the user model*
- **User subscriptions in Dashboard:**
  - *Hyrax::SubscribedPublicationsSearchBuilder < Hyrax::SearchBuilder*
    - Publications filtered subscribed ids for current user
- **Mailing lists for publications:**
  - *MailGun API and PublicationActor < Hyrax::Actors::BaseActor*
    - Create/Updated mailing list from publication actor
  - AWS CloudFront serves download requests to s3
- **Emails for new releases of publications:**
  - *MailGun API and ReleaseActor < Hyrax::Actors::BaseActor*
    - Send email for release create action

Performance Improvements

- Turn off Solr Suggest

- Change Parent Child RelationShip

- Update Publication indexing to get Releases from Solr

Times with Solr Suggest ON

Time to create and add one work to a collection

Flattens out when Solr Suggest is turned OFF. See also samvera/active_fedora#1311.

Solr Suggest Solution

- Times flatten out when Solr Suggest is turned OFF.   Turn off in...
    - solr/conf/solrconfig.xml
    - solr/config/solrconfig.xml
    - solr/conf/schema.xml
    - solr/config/schema.xml

- @awead identified and documented

- See also in [samvera/active_fedora#1311](#)

Reversing the Publication to Release Relationship

Add the reversed relationship in the ReleaseModel...

```
indirectly_contains :member_of_publication,
                    has_member_relation: ::RDF::Vocab::BF2.partOf,
                    inserted_content_relation: ::RDF::Vocab::ORE.proxyFor,
                    class_name: 'ActiveFedora::Base',
                    through: 'ActiveFedora::Aggregation::Proxy',
                    foreign_key: :target,
                    type_validator:
Hydra::PCDM::Validators::PCDMObjectValidator
```

Update places using the relationship

- Remove references to old relationships which look like...

    - `release.in_works_ids = [publication.id]`

- Replaced with things that look like...

    - `release.member_of_publication = [publication]`

Search builders to efficiently find members

- Create PublicationReleasesSearchBuilder which gets list of releases from Solr instead of Fedora

- Added PublicationMemberService which provides controllers with access to PublicationReleasesSearchBuilder

Other Release changes...

- Helper methods in Release model that gets the related Publication

- Indexer that uses publication info in the release solr doc

- ReleaseActor which gets information from the publication AND now has to add the release to the publication

  - `env.curation_concern.member_of_publication = [publication]`

Other Publication changes...

- Update PublicationPresenter to get the list of releases using membership service which gets releases from solr

- Removed unused helper methods in PublicationController
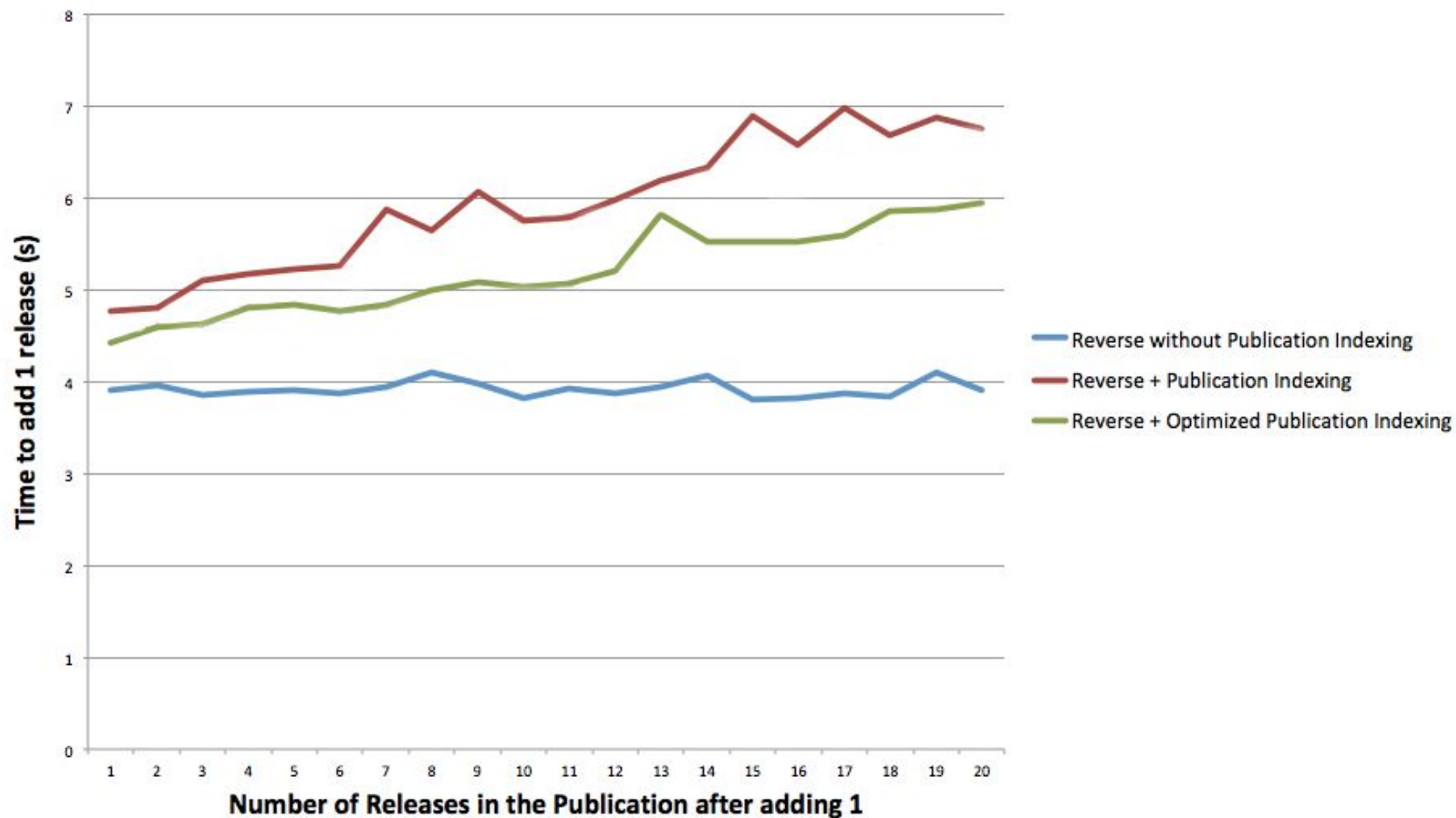
Update Publication Indexing

Original: fetch from Fedora with non-reversed relationship

```
publication.members
```

Attempt 1: looks like it is fetching from solr, but where really fetches from Fedora

```
def releases(solr_doc)
  @release_members ||=
    ActiveFedora::Base.where("publication_id_ssi:#{solr_doc[:id]}").to_a
end
```

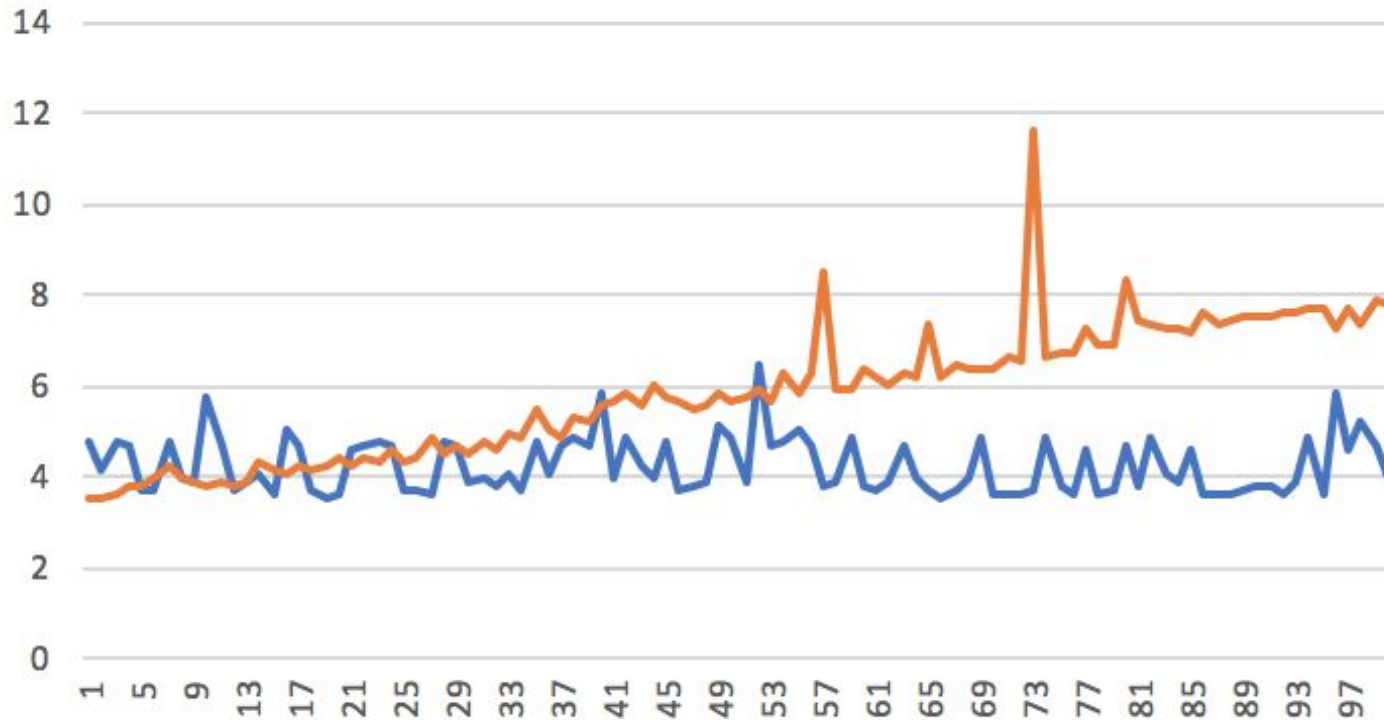**Time to add 1 release in a publication**

Update Publication Indexing (Continued)

Attempt 2: Actually gets from solr by directly using SolrService query

```
def releases(solr_doc)
  @release_members ||=

ActiveFedora::SolrService.query("publication_id_ssi:#{solr_doc[:id]}",
      fq: ["has_model_ssim:Release"], fl: ['release_datetime_ssi'],
      rows: 10000, sort: 'release_datetime_dtsi desc')
  end
```

Time to add a release to a publication

Orange line: ActiveFedora::Base.where   (BEFORE)
Blue line: ActiveFedora::SolrService.query   (AFTER)

Hyrax Plus Deltas

Plus	Delta
Community	Actor Stack
Search Builders	Jobs not Idempotent *(AttachFilesToWorkJob)*
Indexers	Fits
Module #Prepend	Indexing
Rails	ActiveFedora

Thank you