

RePEc Workshop

Federal Reserve Bank of St. Louis

12-13 October 2017

Notes taken by Robert Parks, reworked by Christian Zimmermann

Present physically: Jose Barrueco (JB), Kit Baum (KB), Daniel Eubanks (episodically), Thomas Krichel (TK), Areerat Kichkha (except last pm), Jan Hoeffler (JH, except last pm), Christian Mongeau, (CM) Mahdi Moqri (MM, second day), Bob Parks (BP), Genevieve Podleski (GP), Soledad Zignago (SZ), Christian Zimmermann (CZ)

Occasionally present though Skype: Bernardo Batiz-Lazo (BB), Sune Karlsson (SK), Sergey Parinov (SP), Volker Schallehn (VS), Jan Weiland (JW).

Skype communication was less than perfect, especially for listening in.

Lightning round introductions

(one minute per person, max)

CZ -- involved very early, runs various services, your host.

KB-- involved since very early on, Boston College one of the first archives

DH – here to learn something – worked with Wikipedia data

CM – Italy – new to RePEc – was looking to use RePEc data

SZ – Banque de France – new to RePEc – Big fan

GP – Fed librarian, specialized in metadata

Areerat K. – Lindenwood university– newbie

BP – Involved from the start, used to host a lot of services

TK – founder

JB – In it since the beginning – citation index – librarian

SK – econpapers logec

JW – ZBW

Others by CZ – MM, JK and Dan will join later

Project status updates.

BRIEF updates, including what work goes into the projects. There are potential volunteers in the audience eager to learn about opportunities. Points to hit: what the project is about, current status, what is ahead, pain points, help needed.

Metadata acquisition and overall infrastructure

KB: RePEc email, support for new archives

Answer to repec@repec.org

provide new archive codes and support for new maintainers (CZ acts as substitute). Many new codes never get implemented, sometimes requested again, even for archives already in place.

Instructions are sometimes not sufficient. They date from the 1990's where people have different skills. Need screen shots. Archives may have the data ready but difficulties porting to ReDIF. Many maintainers have little technical proficiency. TK offers help (for a fee?).

Some cannot set up archive, sent over to RePEc Input Service.

Some people are told to take over an archive, have no idea what is involved as predecessor left without instructions.

Action items: CZ reworks instructions with screen shots. CZ writes separate instructions for people taking over archive. TK provides a way to offer help to maintainers (for a fee?)

SK: syntax and link checker, mirroring

checker – straightforward, automatically updated – checks eg wrong archive code cryptic error message sometimes but running quite well

link checker runs on mostly everything – once a week – some issues: big archives – checks in parallel – schedule alphabetically – some servers hit too hard – eg. 5000 checks – some complain about the load – scheduling could be improved.

Not checking journals, do not check some archives

not checking the journals – big publishers OK (well organized) but small pubs not so good

KB Wiley – their links stink – wants to get to DOI – is there a way to selectively do journals link checking -

SK – can be done – took the name – will work on it

What to be done with the errors – sends a lot of email but nothing happening – Did you check the page? – have not been mirrored – issue with that -

CZ – every month maintainers get an email – then they write and say what's up. They do not read emails or go to check page

SK – mirroring – same issues as KB – close to 400 with mirror problems – general

problem with maintainers leaving, new guys don't know what is what, contacting problems, send emails after 7 days and then 30 days – SK follows up with email - a lot of email addresses bounce.

Better approach is to find registered authors at the institution because they care, but this is work.

Action items: SK, KB contact Wiley for link checking. SK, CZ find a way to script contacting authors from institutions with abandoned archives

TK: DNS

names are managed on a server

ssh access – custom build script – rebuild

TK is the only maintainer!!!! Needs backup

needs BACKUP

Action item: TK finds a backup

CZ: all.repec.org, RePEc Input Service

all.repec.org – central repository – any service using the data starts there, indicates where the meta data is (archive templates)

it used to be rather openly accessible, but was put behind web authenticated form. Not yet tested for the new .redif extension, Some archives wget does not work but remi does

KB: problem when institution URLs change, need to catch those and work with maintainers.

RePEc Input Service – for all institutions that cannot host a RePEc archive, especially central banks

about 100 series listed now

works OK – code from SK

only issue: a journal emailed all authors with password to enter stuff

VS: MPRA

2006 online with Ekkehard Schlicht

2015 editor and chief Joachim Winter-

is for people who want to distribute their work but do not have a local RePEc archive

50 editors check for metadata accuracy, 47 languages

41,200 pubs

recent switch from http to https

user registrations 20,867 latest about 1400, seem to trend down

want to convert to more responsive web design
include ORCID-ID / RePEc short-ID
offer DOI's ??

CZ – could enter RePEc short-ID – some archives do that but make errors that are costly to fix – a bit iffy – need to check that is the correct record – this could work by logging in to MPRA through OpenID with RePEc short-ID

JW: EconStor

no big news

intro – econstor disc. Econ archive – dspace software 5.4

148000 pdfs – 46000 items are fetched and fed into RePEc (avoid duplicates already indexed)

cover part institutional services – 200 institutions, mostly German – few others, Austria.

logec ranking – ours is 10 download, 5 for usage last 12 months

still acquiring more series, looking at those with long-pending archive code requests.

Cherry-picking there, as this is not without cost.

KB (+JB, CZ): archive hosting for publishers

KB started hosting journal – screen scraping – Econlit back issues

gradually convinced publishers to send meta data converted to rdf by perl scripts
many publishers have taken the archives back in-house – in their interests to do it
over time more successful

KB still hosts Taylor, OUP, Wiley, Blackwell

Springer – hard headed guy to morally opposed to provide meta data – Gunther Eichhorn retired – his replacement is better – benefit – repatriated their archive – KB offers occasional tech support but now off the plate

Blackwell – Wiley – increasing pressure on them – delay and unfrequent updates – corporate decision not to do it – now much more cooperative – yearly update will be too long – authors complain – Wiley now coming around -

JB – hosts Elsevier – they provide XML, JB converts to ReDIF – talking with them to produce ReDIF themselves – they do not want to host the archive – if we lose Elsevier, 60% of the citation data is lost...

CZ hosts Springer Palgrave on an interim basis, also INFORMS,

also AgEconSearch out of Minneapolis – latter is moving to different platform – they agreed to provide ReDIF – soon online

Dissemination tools

CZ: IDEAS

Covers RePEc in many ways, maybe too many bells and whistles.

Will soon be moving to Amazon cloud, like all services hosted at St. Louis Fed. This will provide more flexibility.

Got some \$ from the bank to have some consultant look at user experience. Indeed, the site was redesigned in 2012, still has antique feel. But maybe users do not care about looks, only content.

Most of site is scripted, but there is some manual work: listing by series (because of inconsistencies in metadata). Site updates every night, not in the most efficient way.

Sometimes takes a full day, when big archives update. The search engine seems to be at capacity, need to look for options.

SK: EconPapers

EconPapers much like IDEAS but with fewer bells

pretty good search engine (Solr)

updates everything every day

takes 6 to 8 hours after mirroring

response could be improved

main problem – SK does not have much time working on it but it is very stable -

econpapers is an odd ball – runs on windows server – potentially be moved to something else – Solr runs on tomcat, the site itself on IIS with very little IIS-specific.

TK: NEP

Marco Novarese is tasked with managing editors and opening new reports. TK does the technical side.

TK – preselects issue content using statistical learning methods

A lot of work – 2016 completely redid the stat learning procedures, now much more efficient

2016 – queuing system for papers – automatic production of issue

the problem was issue size varied a lot, due to large archives dumping a lot at a time.

Now have roughly same number each week

trying to develop bio med news on the NEP model.

need more reports – now too broad – thinking about smaller, narrower reports

Action item: Marco Novarese should identify where coverage is lacking and where it is too broad and recruit new editors.

SP: Socionet

17 years ago started as a Russian site for RePEc. Now a different system – multidisciplinary, 18 others, physics, chemistry

most users from China more than 50%
small budget – about \$800 for the group of 6
in the next session, some new services
CZ traffic increased significantly a couple of years back – Is this due to other fields -
CZ LogEc gets only econ stuff – could be useful to know that physicists stumble into econ
and hence the reason for the increase. SP: it is simply Chinese traffic.

Tools supplementing metadata

CZ: RePEc Author Service, EDIRC
RePEc Author Service - authentication, citation, registration
CZ took on in 2005
TK founded in 1999 – first author system like ORCID
rewrite 2002 – CZ then working on it
CZ still burdened with it – needs constant monitoring – authors mistakenly attribute
things they did not author hence daily checking of logs
Fed paid some for a rewrite of software, now more secure and easier to review logs
other issue is spam – people sign up with homepage that sell socks and other nasty stuff.
Most spam sessions hang on the first registration page – now occasionally spammer gets
confirm email. Delete those, but may have missed some
software is very heavy – most impossible software to install -
takes a week to build – most from citation stuff
Server times out for unknown reasons (not load).
Quite a few authors told to sign up and then do nothing. They ignore the emails, profiles
become stale.
Major burden maintaining addresses – 2% bad emails – remarkable but is a lot of work.
NEP uses changed emails to update mailing lists. Also, take care of deceased authors:
they need to be flagged and may have new works to add.
Maintenance – wants to share but not fun thing
KB – archive maintainers might work short IDs into the papers
CZ This generates more mistakes than I want

EDIRC

Directory of institutions used by author service
currently 13,000 institutions
new institution suggestions come mainly from author service but still many missing
Problem: 15% links bad, and some links work but are not the correct site. Fixing that is
difficult for some countries: China, Korea, Japan, Russia... use Google translate but too
slow
The site is somewhat neglected because of lack of time
Need to add location data to all records, add more institution categories

French are particular problem – centers in several univ. while data is organized as a tree.

SK: LogEc

LogEc tries to capture usage stat for RePEc services

Participating services send logs which are then processed on LogEc server, inefficiently, to produce preliminary stats -

a big problem is spurious access – robots cleaned, cases of author systematic downloads automated weeding.

Final step – manual – look at papers that have suspicious stats. CZ and SK do for EconPapers and IDEAS logs. CZ much of the manual look up – ½ day per month, all manual greping. Weeds out authors assigning papers to large classes, for example. LogEc would benefit from a complete reworking – like a man-year. SK would like a reworking – better automation – weeding needs to be done locally. It helps to have complete view of logs.

Some RePEc services do not upload logs, as they cannot reveal IP addresses hashing would not work. Big miss is ZBW, INOMICS also uses the RePEc data but is little used

JB: CitEc

Second citation index on Internet, started in 2001.

29,000,000 references

Initially mostly download pdf, then find references. Now moving to metadata provided by archives: big publishers have the XML and can do it. Also, authors can add citations on CitEc site. RePEc Author Service also adds citations, but no one knows how.

CZ Some people complain that citations are missing. This affects the credibility of the rankings using that data. About 1 mio articles are not treated. Some are in Cyrillic, but unclear where the gap is.

JB Problems: citations are gone if handles change. One publisher changed handles three times in a year.

Also: intermediate pages preventing direct access to pdf. Eprints and Dspace sites we can deal with, if we know about them. Others are custom. See CitEc site at <http://citec.repec.org/topcoverage.html> to find whether a series could be captured properly.

SP: CyrCitEc, SocioRePEc

CyrCitEc – Sponsored by RANEPa for 2 years.

extract content from pdf – good way to extract

extract sentences – try to anal citation content provide CitEc data in for cyrillic content.

TK – going to work on conversion from pdf – Json – font changes – main thing to context of the citation – potential pos or neg reference – parser for separate Russian fonts – build training models to better detect citations

how to visualize – sociometric data in references
SocioRePEc – RePEc information system like IDEAS

CZ: Genealogy, Biblio, EconAcademics

Genealogy – way to get authors more engaged -- academic family tree – who did you advise, where, what year

about 1/4 of authors are in – good start

used for creating a ranking of grad programs

some departments register students

minimal maintenance

Biblio – have editors determine most important papers of a field, then subfields which they can assign to others – not reached critical mass – need to recruit – have about 50 editors – about 125 topics – unequal treatment across fields – encouraged NEP editors to no avail

having a syllabus in a field course can already be a start

There will be a recruitment effort by the end of the year

EconAcademics – assemble blog posts relevant to economic research– scans blogs for links to RePEc – link back to post from IDEAS

gets one to 10 posts a day

blogging is not as popular as was

some good blogs – EconAcademics help get the blogs known

DH – what about including tweets?

CZ – Twitter throttles and need a list of people

Note that list of blogs needs maintenance: moved blogs, dead ones, add more.

Note also: if a RePEc page can be found on Wikipedia (using its API) – IDEAS links back as well.

TK: CollEc, "rogue" publishers

CollEc – dormant – finds who is the most central author

closeness – calculates all 33,000 authors paths – $33,000 * 33,000 / 2$ – summarize – big compute power – shortest paths for ones that update – search engine – 2nd server helps full path data exported to some other guy that was supposed to write a paper – not fully backed up -

issue of sustainability – getting too big - has a sponsor and a machine

CZ – it would be great to have a graphic representation of network

TK could do by subfield, which is a smaller problem to solve.

somebody pays TK to do it, it will be done

rogue publishers – have a committee – dormant -

idea - committee assessing rogue, predatory, dishonest, deceptive practice publishers,

bunch of disreputable publishers got a listing in RePEc, giving credibility - sanctions? – list of bad guys – wipe them from RePEc – if they are out of our store we cannot show they are bad -

TK would like to write constitution -

KB was looking up Jeffrey Beall's list – should we index if you are on Beall's list, why are you going after us.

TK we need a policy

KB – when we get a request for an archive code, what sort of review should we make?

Hold on with the request for suspicious ones? Many codes never get implemented. Wait for an archive to be live to act?

Action item: TK revive this

CZ: Plagiarism Committee

Deals with suspected plagiarism for RePEc-indexed items – on hold until CZ finds time – several cases waiting – labor intensive – 30 cases adjudicated – IDEAS flags plagiarism on abstract pages. Some people have lawyered up – lawyer sent stuff – no one bailed – lawyer emailed the admin of the committee – no one reacted – remarkable how steadfast – lawyer disappeared.

Critical: procedures are well specified – 2/3 majority – a nonvoter can kill it. Thus difficult to get some past the committee.

Tools that supplement dissemination

CZ: myIDEAS, ranking, API, monthly mailing

MyIDEAS – personal bibliography tool as you navigate IDEAS. Also allows to track additions of series, authors, JEL codes and search keywords.

about 2,500 users

once to the cloud can do weekly digest email. Currently, you need to visit site to get updates.

planning to add user notes to abstract pages, via myIDEAS – some way to leave public annotations for authors and readers

Rankings – many people think this is all what RePEc is about – unfortunate focus on the rankings – not happy because rankings are noisy, reputation impact

Important though, rankings have been a big driver – got many to register or to get publisher participating.

API – allows to have access to RePEc data – too difficult to go the regular route, especially if you need just a slice of the data – 30 users for the API – researchers, institutions to evaluate their economists, and some abstract services - Proquest,

Altmetrics, etc. – advantage of API CZ does only provides URL to IDEAS (or EconPapers if requested) rather than to the pdf

Monthly mailing – CZ does to maintainers, authors, editors

2 times a year reminder to registered without pubs and EDIRC institutions.

Goal: get more loyalty, reminders that they have stuff with RePEc

Many people are waiting for the statistics, then others who do not want the email – spam flags

Requires work: dealing with bounces, flooded with vacation emails, all the captchas - dealing with responses – standard replies help in most cases. Less since FAQs have been reworked.

RePEc Blog – Need contributions. Ideas are good, too!

TK thinks we need to host blog – KB hosted – then got hacked

CZ with wordpress you need to be on top of patching

Closing some projects?

No suggestions

Opening new projects?

Gateways from other formats/publishing platforms to ReDIF

CZ we should have plugins for some of the software

TK – possible – using XSLT cumbersome

unique identifier – easy – the rest series, archive – must be managed within the plugin

TK the transformation not just the record itself, but need other documents

KB – some archives use DOI – use them

TK – DOI not as stable as people think – KB but better than most else

CZ should be using volume, issue, page number for unique identifier – but some physics journals have more than one article on a page

KB – no numbers on the pages in some new (internet) pages

CZ plugins should be truly plugins: easy install

CZ creating a plugin needs people familiar with software and our metadata

TK VS knows Eprints -

KB many OJS want to put into RePEc -

TK meta archive soak up Dspace

CZ meta archive don't need plugin just need central script

TK standardization in Dspace? Jose no.

TK each archive needs some specialties

JB– LSE has a custom script for Dspace – could ask them

JB email LSE guy, CZ takes Eprint. Who takes OJS?

CZ no converter – TK copy and paste CZ yes we need plug ins

TK no agreement on XML formats – Now JATS data – most journals know – but complicated

KB had inquires from ? JATS data – they were concerned wanted to give JATS data so citations could be pulled – TK JATS can do citations

KB claim is JATS is the standard – used by pro publishing community

CZ OJS one journal at a time –

CZ converter from JATS and bank on every platform has JATS output

NIH requirement ???

TK OAI gateway – XML – NEP is based on the XML data

TK worried about JATS to ReDIF

CZ JATS – need to track, TK done by NIH, TK we don't need all of it but we could scrape it

GP – knows something about JATS -

CZ need converters to ReDIF – data needs to be in public files – fundamental change if accept in other than ReDIF

GP – building pub platforms – put them in directories

Jose – we need has expertise in OJS not JATS

CZ we could identify the OJS archives

Action items: JB contacts LSE regarding Dspace. CZ contacts VS regarding Eprints. CZ contacts archives on OJS to see whether someone has a script or plugin.

ArchEc

TK: Fundamental – given the decline in usage, as a community, we are exhausting the end of meta data – we need archive function or publishing function – ought to be around working papers -

econ has the working paper culture

the core operation ought to be moving to an archive operation for working papers

This relates the beginning with EconWPA centralized versus decentralized. Decentralized is not serving us well, go centralized.

ArchEc adds an archival function – paper still there even if the institution is gone. Over past few months conceptual progress – have a worthy system. First operate as a dark archive, then open it up to use.

Example – NEP – all that it did was go to the URL – it would link to a stored copy of the paper on ArchEc

This needs funding, and you get this only if you have a concept. TK did not have one for a long time, but has one now. Now it is easy.

DH what about a distributed archive?

KB – first, what is problem is this solving. Yes links go bad, if a WP is 5 years old and never published, what is the harm

TK his first paper is not published but has more cites

KB no doubt that some WP are influential

KB is making copies of every WP something we should do

TK argue building the architecture – then encourage usage

TK the problem – logec is telling us usage is falling. We need a new concept. ArchEc makes RePEc central again.

SK If measure all usage take away all other copies – that seems dangerous

TK the working paper culture in computer science is gone. We can preserve it for Econ.

CZ plenty of pre-print archives in other fields are appearing
CZ – CitEc gathers papers – already dark archive
TK not visible, not integrating full text and meta data
CZ if you want to make it public, you have a copyright issue. See how it worked for Google Books
CZ – as long as dark, no issue but if public then some author may not want it visible
CZ other areas all face funding issues because they are costly to run: they host pdfs
CZ we are really cheap to run, we link to pdfs
DH if no one uses it, no problems
CZ to Jose – do you have a dark archive YES
JB but not all since the beginning
The pdf files do not have the meta file
CZ – the emergency is to have the files
CZ – make sure we have dark archive mass – then sit down and think about putting them together – CZ has backups of ReDIF files if needed, as long as they have not been overwritten.
SK one thing is archiving – basically good, useful, different ways in the future – figure out how to do it – thorny issue is what can we do with the stuff – implicit agreement is we have meta data and links – if someone wants to get rid of their paper they can now – part of the agreement –
TK should have the standards, after 5 years think about how to use

ReDIF-Perl rewrite

TK This is software to read ReDIF, used by all RePEc services – put in structure to access elements – written years ago – believe needs to be rewritten, published, documented, so others can contribute
SK most error messages on the checker come out of ReDIF-Perl – but have others that come from other
TK – funding for ReDIF-Perl – ArchEc is related to templates – ReDIF-Perl creates the templates – work ought to be jointly done – along with policy statements – the code and policy decisions are getting stale -
TK more sound basis for collaboration
CZ is ReDIF-Perl inefficient? code that has errors?
TK we still have perl 4 problems – (mirror by SK) –
SK not too difficult to replace the mirror
SK ReDIF-perl is difficult to maintain, hard to maintain, other things are higher priority – it is not broken -
CZ need better error messages – other aspect: it could treat poor character encoding
SK – spent some time on poor encoding – if the file is unicode and the archive says not

Unicode, that is bad

TK lots of lines special to single case – fl ligature -

MM why not put it on GitHub?

CZ suggestion to put ReDIF-Perl up good one – better error messages would be great and anyone could do that – but now cannot do it

TK better comments, guidelines, no one wants to do

CM where is the code – answer the repec code archive

KB – Sune issue of encoding – go to check page more clear indications that they should worry about encoding – informational message

SK that is difficult – UTF8 is valid win 1252 – hard to detect those issues – warn about characters invalid – UTF8 or win 1252 – most of the warnings do not come from bad encodings but rather from copy paste – UTF8 copied latin 1 -

TK – data is in UTF8 is forced into something else – double encoding – infamous a~

SK that happens – UTF8 looks like latin 1 -

TK – reads correct latin1 with the continuation byte but is incorrect

Action items: MM opens a RePEc account on Github. TK puts ReDIF-Perl there, other contributed scripts from RePEc:rpc:script as well. SK look into feasibility of encoding warnings in checker.

Reassessment of RePEc usage conditions

TK some of that stuff – move to someone else and come back

TK annoyed at usage conditions – where are they displayed

display – usage display from EconPapers

TK – written when feared SSRN and some others would use RePEc data without attribution

TK – data is not open data – restrictions – point c) must contribute

TK entire statement should go

TK replace with RePEc data belongs to the people who provide -

we don't assert any rights over it.

TK we believe that the data contributed wanted to be widely disseminated

New statement was hashed out and is online at <http://repec.org/docs/RePEcDataUse.html>

TK a) the Board must approve b) maybe need a new Secretary of the Board

Renewal of ReDIF and Guildford protocol

Guildford protocol – rules how ReDIF is written to server

TK did some work recently

TK has a new version, bring up tomorrow

Action item: TK make the new version live.

Expanding in new directions?

CitEc use of CrossRef, OpenCites

JB: CZ suggested – new movement distributing reference data in OA – initiative for open citations – reference data

JB All the movements has been generated by the big publishers – for us we have this data already in CitEc – not sure that CrossRef will add value – small publishers do not contribute to CrossRef

second – all this data is heavily DOI – article with DOI go to CrossRef – only to access – for us the limitation is that use of DOI is suggested not mandatory, thus little present there is OpenCite has 5 million references but we have 29 million

CZ data quality – is big pub good quality data - Could CrossRef help cover older material.

JB were are OK

CZ looking for opportunities for data – another source is Sci-Hub that stores pdfs from publishers – could be a source of references but is considered illegal. Problematic CZ – wants better data – IDEAS uses references from working paper for the journal article -

TK we are being copied were pioneers but not now

JB – problem with the small publishers and working papers

having the archive of pdfs would help – departments that hide or gateway the journals

CZ we have 1.5 million articles in IDEAS, but 0.5 million with references, gap of 1 million articles which are not with references

Action item: JB identify where we have a lack of coverage among journals.

Discussing archives which have problems

SK – freeze the archive – remove ones that do not work and small ones integrate with the link checker

More on this later.

Peer review

CZ We are sometimes asked why have not done a journal or something

CZ RePEc has always been focused on WPs, which are not usually peer reviewed

Why not some sort of peer review through RePEc? People have suggested that. Are they thinking about yet another journal

or is it a rating system – ResearchGate is trying to do

worried about Facebook features – CZ not a healthy thing to do

TK NEP editors do a ‘rating’ – can move papers up and down

TK has tracked the NEP moved up or down – Marco did something like that – are editors picking winners

CZ – NEP move may have some effect on downloads

Can we do something more explicit – there are journal management software – easy to do

RePEc has a positive reputation – not commercial – we have the research interest in mind

Should we get into the business of a journal -

TK – never had a strategy – never found a structure to compete with existing journals –

TK does not believe we can do a journal – but there are other ways beyond NEP reports which is a first stage

TK maybe based on NEP or parallel to NEP -

TK – issue with journal cheap to do -

TK thinks increase in commercial -

KB AEA 5 journals

CZ Swiss will have Springer take over

CZ \$75 to put up a PDF – publishers blind the scholarly societies

RP – Bill Barnett example of going to Cambridge

JH – two new replication journals not working – do not attract good papers

TK – potential is business case, within the existing is a stone unturned – a bit of tech wizard – create some effect

CZ – proposal – How does publication process work now – submit to journal – highest possible – go down the ladder to the right spot Want to reverse – start with NEP report –

Journal of labor supply elasticity – very esoteric – then promoted – to Journal of

Empirical Labor – then if good enough – Journal of Labor then Journal of Economics

TK – avoid the name journal -

CZ – Call it RePEc Peer Review: XXX.

NEP-DGE blog, already selects a paper per week,

GP could also be part of Biblio

KB if took 90 NEP reports into 10 or 8 areas – which are the best this month for this area, a board review

CZ would the NEP editors put in the effort

TK but a number of people should be fired – Marco and others arguing to keep them

TK illustrates with some guy – could not fire him due to his silly comments – TK thinks damage to our reputation – TK relates bad editors

GP could you recruit someone to blog about top papers

KB that person would be chair of the board

KB do you have some deadbeats in the 7 or 8 that a ‘chair’ is over

CZ – summary – do not talk about Journals -

CZ – journals will be less important in the future as article-level citations will be more used, rather than impact factor.

TK – straight jacket of print – as print hard copy disappears chance to change – the future

is bright – spreading out

KB – ArchEc

TK – publishing model so it can be run -

KB you can already write with appendix online to do the stuff

DH someone did it – fiddle

CZ run my code – project – upload the code – online change the code and rerun – Vicki Stodden – runmycode.org

CZ maybe make a proposal to restructure NEP into groups with blogging, they can also organize themselves to create new reports

TK is technical and wants to stand away from running it

TK Marco needs to do that

DH – any person can rate a paper -

CZ – only rating we have so far is on RePEc Biblio – fired an editor thanks to it

Action item: Marco needs to explore grouping NEP reports, each group managed by a board that determines whether new reports should be opened, and that runs a “blog” featuring cool new papers.

Fail-over for IT resources

CZ: Does every project have a plan if IT resources are failing? We will be asking each project manager to explain how this is handled (or not). We will identify gaps and try to address them.

What happens if something goes wrong?

At UConn, CZ had two boxes with no backup. CZ was motivated by Fed due to support – bank would provide and administer the boxes, they had plans to replace, upgrade. Move to the cloud adds more resilience, about an hour to switch to larger instance, backup solutions, thinking about if part of the cloud goes down, another part takes over. For stuff that CZ has, there is a plan.

SK stuff – fail-over nonexistent, EconPapers and LogEc on virtual server by the university – if he gets another box he can move quickly but he does not have a box.

TK – all boxes are at Hertzner. 5 machines. sponsors for all of them

get sponsors to upgrade every so often. Hertzner never had a problem – never had an issue. Each machine has a different policy. Hertzner allows to do anything on the box. very reliable – if something goes wrong, usually software

in the existing machines, the data is held twice – disks go bad, they put in another disk – fairly reliable – every box mirrors some other

IKE2 and Katsy? Are large 32 terabytes and mirror the others, but some data not copied.

Maybe one or two more boxes could help but no more than that

A couple more sponsors might be nice but otherwise good.

Only issue with Hetzner has been in 2000 when Hetzner dismantled a box due to lack of

labeling on the physical machine

Would be good to have backup DNS server in US – all current DNS is on Hetzner

CZ how stable are the sponsors?

TK oldest sponsor since early 2000s – do not have full admin access – have to contact.
full access for other machines

TK other day – had issue -

CZ if a sponsor drops

TK need to distribute the load to other machines

IKE2 is the mirror and would be hard to replace – sponsored – Auckland no longer paid –
it was a scramble to move the machine -

TK usually takes 3 to 4 days to move

**NEED LIST OF SPONSORS, MACHINES, SERVICES FROM EACH IN CASE A
BUS HITS SOMEONE OR A SPONSOR**

MM \$500 per year for a really good server

MM loves the cloud

If Hetzner went down, take time

Need to have a sponsor requires a Service which is visibly sponsored

Need to have the logo somewhere – but something acknowledgable

CZ at UConn – need a sponsor from academic societies – response was OK but want
special services for their members

CZ only success one society for which he was secretary

CZ very difficult to find sponsors

CZ provides services to NBER, which sponsors a Hertzner box

TK everything is about the same except for ColLEc -

CZ can get sponsors but TK cannot have root access

TK cannot do stuff without root access

KB – what about ArchEc – TK just storage space

CZ – it looks like we are largely resilient except for SK's stuff

TK – SK is running on Windows -

TK CZ – we could at least back up SK's data

SK (later addendum): EconPapers and LogEc are backed up on the SwoPEc server at the
Stockholm School of Economics and a desktop at home.

CZ MPRA – small piggy back on the library server – probably good, plus full support
from the administration

KB repec home page, threat is hacking, otherwise OK

KB other thing is stored at BC on ftp site – CZ says have backed up

KB could get a copy,

TK other pages, KB used to be hosted by KB, now on an IT box virtual machine

Action items: TK see with SK about backing up his stuff. Everyone document on the RePEc wiki (see below) infrastructures and sponsors.

Fail-over for volunteers: substitutes, succession.

Does every project have more than one person who can take care of things in emergency? Can this person eventually take over permanently?

KB no redundancy of TK even though the boxes have redundancy

No backup of JB, or CZ, or SK

TK and JB – no human backup

TK is trying to commercialize a NEP clone in biomed

CZ case – CZ is completely alone – cannot allow outsiders into the system

hope with the cloud move that there could be outside access

CZ system admins at Fed have access and have some knowledge

there are internal wiki pages – procedures – e.g., monthly mailing

all is detailed – there is some fall back – could be done

there is some documentation and some knowledge and if necessary could act on it

CZ document enough that someone who knows perl could do it

JB ???

TK – CitEc code very difficult

TK – perl is declining – would like to replace with something that is more future proof – and more popular in current generation

TK would rewrite ernad in C

Author service was originally in C -

DH – python – 3 is critical mass – python or javascript

MM – open source software, as long as documented – e.g., converter from perl to python – javascript is limited

MM as long as well documented -

KB most important – take the worst and write the documentation

KB shouldn't all be on Github or somewhere

RP buy voice recorders, record voice comments

MM documentation separates the human and the code

JB – we are the same group for 20 years

TK commercialization

JB – mostly maintenance rather than exciting new development

CZ – more and more people are interested in RePEc data for research

CZ some want new things, e.g., history of affiliations

KB author stuff could be history

KB – most useful is RePEc data linked to other data

CZ excitement for new things but need new services

KB expansion of the service – enhancement – eg RAS someone can look up how individuals move around
DH expand to other fields – potential from other fields – some complimentary
DH a lot of open source projects – way to deal with it
CZ – put stuff on Github and get people interested
MM Github tracks changes in code – we put all code on Github – people record issues, tracking the code, any policy against Github
CZ At Fed we used a private github – cost a lot of money – moved over to GitLab – thinking of offering it to our economists for version control for LaTeX writing – based on Git -
MM well supported, well hosted
CZ DNS anyone else but TK yes JB
CZ is there a wiki on Github – could be used to explain who does what where
CZ volunteers page
MM converted perl to python – but without git
RP – high school projects could contribute through Github
TK has tried for sponsor
TK – got grant of WoPEc and for RAS
TK – sponsorship of machines but grants to make progress – e.g., ArchEc
CZ – Sloan foundation can be interested -
TK can supplement what he has for a grant application
TK – two parts for ArchEc -
CZ combine with CitEc to get support from Sloan
CZ Github part of grant
CM – TK said we are too geeky – social network – share what you do
TK – to CM not easy to understand what we are doing
CM – important to spread – if people don't know, how can they help?
TK – NEP – general editor has instructions – gives some info about the internals
CZ in some ways we are open but the rest of closed – the code, functioning is all closed – high cost to enter our world
TK – has concept of an openlib – show me the code that generates the page
TK – do something similar – maybe put code on Github and hope that someone documents
TK – CZ – economists are on Github – plenty of coders -
CZ if code is modular then mix languages
TK maybe maybe not

Action items: MM create a Github account (done: <http://github.com/repec-org>) and a wiki (done: <http://wiki.repec.org/>). Then everyone put code there and add stuff to wiki!

Mailing list policy.

TK do not know how many mailing lists

CZ repec-announce is open to anyone, but had problems with some ill-intentioned people subscribing. Then did repec-run, which is not listed

TK – worries about historical trace of discussions

TK – need list of public repec mailing lists

CZ if public then we may get more volunteers

CZ helpful to engage in coding if they can see what is going on

CZ historically somethings do not want publicly visible – e.g., grant proposal

TK repec-help list, repec-archive list

CZ repec-help list people to help with new archives, dead

CZ need to be easily discoverable, Github will help

DH can get code – there is an API -

CZ – can get email from Github

TK discovers broken robots.txt files on list servers, fixes it to make it visible to search engines.

TK should we do something now with the lists

GP – attract people not exactly in the wheelhouse

SZ private blog for RePEc members

CZ to respond must go to email

GP – fewer lists, more content, and publicize

GP make it easier for outsiders to participate

CZ – more general list, and specific who would post digest

CZ want to treat repec-run as a general list, specialized lists like citec-run could post weekly digests

Jose – need list of lists on main page

CZ which lists to list: repec-announce, repec-run,

CM what is purpose of repec-run – what is the general public

CZ repec-help who would help build an archive, but is intended internal among helpers

KB post to repec-run things to work on

CZ mailing list archives are chopped up, impossible to find something

GP – want to see what is going on

CZ bulletin board/forum – any body could ask question – can respond to it – answers are public

KB forum not a bulletin board – anyone can view but only registered can post

GP where do requests for volunteers -

TK rarely see much about NEP on repec-run

TK changes options for repec-run: make it visible (done)

KB list repec-run and repec-announce on the RePEc home page (done)

Social media: can we do more? Should we?

CM: social media is a way to make people aware

maybe get some interested

plenty of social networks – facebook, Twitter, LinkedIn, Github

looked at RePEc services – some have integrated social features – IDEAS has a share button

NEP does not have social features

for most of them it does not make sense

MPRA – makes sense – share paper – have an easy way to share a paper but in current

MPRA does not have a way

could not find a EconPapers share button

SK there is one on the abstract pages

CM improve the sharing of links – Twitter cards – small picture of the abstract

KB thinks that SK does it -

CZ they come from me (NEP Twitter feeds)

CM shows example tweet + image – gets more shares, more recognition

CM thinks not so hard to do to get an image of the abstract

SZ thinks even the logo would do

GP – Twitter code card code – button

CM other thing the button does not work to share – on phone maybe – works on other

CM more and more mobile

CZ getting 20% of hits by phone

MM reads more papers on the phone

Facebook has similar – Twitter card can be used to send to Facebook and LinkedIn

SZ says that is how they do it and GP says already has the code CZ will implement

CM communication point of view Facebook, Twitter, LinkedIn accounts for RePEc

TK singularly unable because no person talks for RePEc

GP tweets by CZ – do not need the official RePEc thing

CZ reason started blogs – visible activity – we do things – refers to blog posts in FAQs –

the blog is effort – needs to be regular – with tweets less work – CZ is not on LinkedIn -

TK – Russia blocks LinkedIn

SZ better if more than one -

GP can automate feed from blog to twitter

CZ blog posts are automated

CZ does not want the trap of top 10 list

?? Disseminate top papers on social media

KB – NEP lists ordered, what would it take to snarf up the top of the NEP posts

TK we could automate in ernad (NEP) – we know which has promoted

KB every week put out a post that was at the top of an ordered NEP list -

CZ – influence of ranking

TK learning is binary -

CZ need to make this clear

TK filter is pre – selection screen by topic – filter is excluding like a duplicate paper – the ordering has no effect on the selected -

KB some scrutiny

TK if it is pushed to the top -

KB only good editors first

GP want a link back – one of the things we put out is the top

CM study say 62% of users get news from social media

CZ but we do not want the general public

CZ – log analysis shows sometimes Facebook or Twitter generate traffic for a papers, but usually it is blogs – even papers mentioned on a comment on a blog post. But in any case, the value is to make RePEc visible, as the Fed tweets do for FRED.

SZ reputation effect – gets in peoples head – can not measure but in terms of reputation
yes

CZ says central banks use twitter for that

KB central bank not just about policy

GP if talking about RePEc present yourself

MM just created a Facebook account

CZ KB tweeted and Marius Ooms retweeted

KB Marius was editor of econometriclinks website

CZ Marius could manage Twitter – health problems – what he is doing is very episodic – the @repec handle is someone is named Repec – inactive since 2010

CZ Marius manages @rep_ec, cannot take it from him – but maybe could have someone else involved – trying to share would be better diplomatically -

CZ Repec is Czech name – also a small village – April fool RePEc blog post sponsoring the soccer team

CM should have Twitter account etc

MM already done, @repec_org

CZ on the NEP Twitter accounts sometimes other than a paper announcement – marketing for RePEc

CZ Twitter does not like bots – occasionally thought to be a bot

CZ even paper titles might trigger the bots

CZ is registered as a developer AP

Action items: CZ reaches out to Marius. MM has created social accounts. CZ do Twitter cards for IDEAS and possibly NEP tweets with help from CM.

Drop in measured use of RePEc services

CZ: loss of traffic -

CZ one decline – Google changed page ranking formula in May 2014 – aggregators took a hit (~-30%)

CZ most users come from Google, land on one page and leave – users do not do more pages

CZ and others – Amazon model – other things interesting

KB Elsevier (ScienceDirect) page says here are other interesting articles

CZ – what is our user – professor, student, central banker, general public. Same with FRED the audience is heterogeneous. Difficult to satisfy everyone

MM google who comes

CZ – result of search – corresponding with scholar google – put stuff in meta tags

MM 400 words to get in Google -

CZ – EconPapers does not do well with mobile device

CZ RePEc data used in more places but not reported. ZBW none of the traffic is reported

TK exactly the point about what Google presents on search for ‘money is memory’ – aggregators are losing and we have to do something else

CZ - we want to go to abstract page -

JB – permanent DOI – need handle – EconPapers has a permanent URI, CitEc is doing it rather than URL

CZ handles change – some journal changed handles

TK we need a central registry for series

CZ SK is maintaining such a list and CZ uses – but RAS does not use

CZ is trying to increase traffic -

TK we need to change from aggregator to archive: ArchEc

GP would compete with libraries and institutions

TK especially the institutions that do not have archives

CZ is RePEc primary or secondary source?

TK we are way ahead because we are distributed

JB we need to start a new service

TK need authorization to do it

CZ nobody opposes a dark archive – CitEc already has

CZ when ready can come out of the dark. Allow publishers to opt-in to have pdfs open.

Default open for defunct institutions.

TK back to argument to ArchEc – and rewrites of perl etc

JB – we are losing archives – and yet growing over time – we need to do something about lost archives.

Action items: CZ investigate 400 word limit for Google indexing.

Under-represented groups: Heterodox? Women? Developing countries? Non-academics?

SZ: First suggestion: NEP reports are thematic – if we want to expand is possible NEP by regions – papers produced by the region not about the region. Similarly, NEP reports could be created by type of institutions: central banks, think tanks, to attract/cover more non-academic authors.

TK could be some automated

KB issue about national affiliations of coauthors

TK needs to be first claimed by author – claim is much after paper is in NEP -

CZ – we have NEP-GER, German papers – by language

TK we can – could also open NEP-XXX for papers unreported after three or four weeks

TK had proposed AEA any claim from author in AEA announcement list

TK we can automate

CZ we should create communities. Geographic NEPs can help. CZ was editing NEP list on central Asian, it was very easy – maybe find someone who could do many lists topical geography

TK so many ways to improve NEP lists

SZ second one – related to female representation – data on female representation in IDEAS – this brings visibility and that RePEc is engaged in diversity Can we go further - look at the selections of NEP editors – how to get more diversity – NEP gender issues is by a man – new NEP with females

TK if we can get new editors e.g., Islamic banking -

CZ NEP-GEN is a recent creation, no female volunteered before. Tried to recruit heterodox, got blank stares at ASSA meetings from the ‘clubs’ when talking to them --

KB – push in England

CZ tried for years, nothing -

How do we get them interested – identify gaps -

TK data is public about which papers are not included – TK did a paper – the more a guy writes the more likely to get in a NEP report

CZ biases the other way – 10 papers from same author only put a few of them

CZ is there a way to automate selection NEP by language

TK yes we need an editor – because the abstract English but paper in other language

TK did NEP-XXX – English abstracts for non English papers

TK particularly Asian papers -

TK would like to have a NEP editor for Russian papers

CZ BIS is running a service that aggregates all papers from central banks – NEP could also do reports by type on institutions: CBs, think tanks, etc.

SZ return to topic – topics – institutions by female heads – female shares -
but the data is flawed
CZ thinks that females register less than the population
GP what could be interesting could start the conversation about the competition of
representation of female
CZ approached CSWEP – talked to the incoming president – replied what the heck is
RePEc, I am not interested
SZ has convinced some to register – the women who are not registered are not well
ranked – instance promote a list of French women economists – Twitter followers –
maybe a ranking 25% best women – so journalists can contact - CZ critical mass -
SZ can do rankings for countries with at least 1000 authors, others just an alphabetical
list.
SZ – finally, we need to work better with other languages – short presentation in many
languages – could have one paragraph about RePEc in several languages

**Action items: CZ: list of best women by country. Marco: look into regional and
language NEP reports. KB: RePEc in a paragraph for RePEc home page. Others
then translate it. SZ: look for NEP editors for institutional and country reports.**

How to handle abandoned archives

CZ: What should we do with the abandoned RePEc archives: – freeze and archive them,
have them just there, delete them – what happens with reawakening
KB had a recent example – suggestion is a ‘diff’ of the our archive and theirs
KB priority – try to keep from disappearing -
TK wants to get funding for ArchEc that can recode ReDIF-Perl
CZ is this a RePEc meeting or ArchEc meeting?
KB in some cases the fix is easy, a url changed from http to https, but the listed contact is
obsolete.
SK (addendum) There are two kinds (see bottom of <https://econpapers.repec.org/check/>).
Inactive archives (353 archives have not added anything in the last year) and archives
with some mirroring issue (377). As far as I know, nothing systematic is done about the
inactive archives (except Czs monthly mailings to maintainers). For archives with
mirroring issues, as automatic e-lain is sent to the archive maintainer after 7 days with
problems, then after 30, 60, 90 days and so on. For the first three, I follow up with
personal emails detailing the problem and what they need to do. If the email to the
archive maintainer bounces, I try with series maintainers, If these are different. Then with
registered authors at the institution or some email address I can find on the web, contact
email, department chair etc. If I have no response, series maintainers etc get added in the
later follow-ups.

Of the 44 archives that developed mirroring issues in August, 11 still have problems. So, about a 75% success rate. The issues could be fairly trivial, like a single link to a .rdf file that gives a 404 error.

“Freezing” archives is a technical issue and pretty straightforward. My opinion is that this should be integrated with the link check to weed out bad full text links and keep the good ones, if any. Related to MM's question in separate email, I have backups going back to August 2003 and I guess others have similar backups.

The hard issue that takes a lot of time and effort is to get these archives back on-line or active again. This is tedious and boring. The email we have will likely bounce or we get no response. Then chasing down other contacts, possibly having to explain what RePEc is and in general getting them motivated to do something about it, To be feasible this needs to be divided up between several people.

Action item (added later): CZ integrates freezing archives in the RePEc:all interface. Frozen archives to be archived on RePEc:all. Can be unfrozen if reawakened.

Varia

CZ: issues of journals with missing issues maybe SK could help with syntax checking

Action item: SK look into journal complete check

New volunteers

CZ new guys in audience: how do we attract new volunteers

CM – who are the old volunteers –

CZ we need people everywhere

GP is there a web page – here is the RePEc universe and all the people

KB RePEc home page – the core team -

CZ volunteer opportunities pages on IDEAS

GP and SZ call for volunteers –

GP what did work – CZ TK really random

MM marketing and social value – the page for volunteers not specific enough – which projects -

JB – home page needs redesign – maybe get volunteers to do that

KB – maybe change to button and pop ups

GP information is more important than the look – GP looked for a year to try to figure out what RePEc is

MM where do I go from the home page to find out what to do or what is what

KB services are different representations of the same data

GP are you an author look at this, if you are an archiver look at this, if you want to search for Then here

DH description of the distributed nature

CZ need a visual picture of what we are

GP people would like to know what is input -

CZ is getting traffic from

GP use RePEc Tools – RePEc services is old terminology

CZ how did you come to RePEc

CM – got interested when knew could access RePEc data – thinking about some paper to do for some project – sent email how to get data

SZ – short term with Banque de France – friend – tell me about ranking – contact for missing citations – understood little about how it worked – Banque de France is a big fan of RePEc because it is well ranked Many institutions are interested for similar reasons due to rankings – think it is a very sexy project – look is not so important but the information is

MM business research data – used the data – most cited authors – hard to find the data – was scraping from all the journals – data has a lot of value but so inaccessible -

TK we have not been aggressive about sharing

DH open source projects – part of it was geographic -

GP not sure about how helpful my story is – thinks there may be similar librarians with skills and interest

GP need authority for authors – RAS does this – recommends to others.

General action item: we need to open up, show what we are doing, use github and wiki to document code and processes. Be present on social media. This will entice others to participate.