

If you are using a mic please use headphones to avoid feedback

Built-in speakers and microphones create echoes, feedback & spoil the session for others

- » Click on the profile icon at the bottom of your screen
  - › Select 'Setup your camera and microphone'



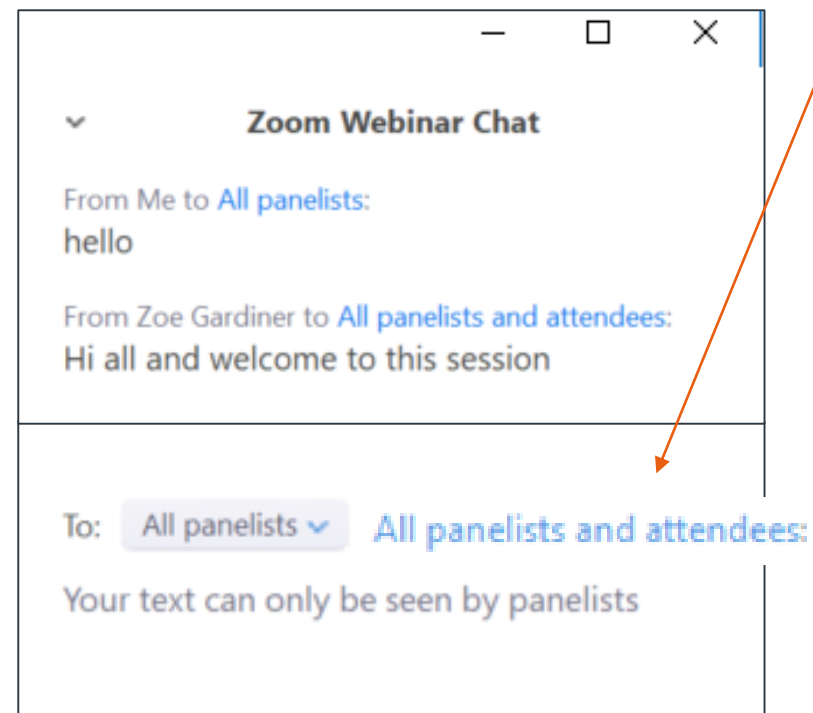
<https://flic.kr/p/oPtKMj>

- » Use the icons at the bottom of your screen to:



- » We will be using the Chat panel to answer any questions

- » In Chat, remember to change the drop down to 'All panellists and attendees', so everyone can see your questions





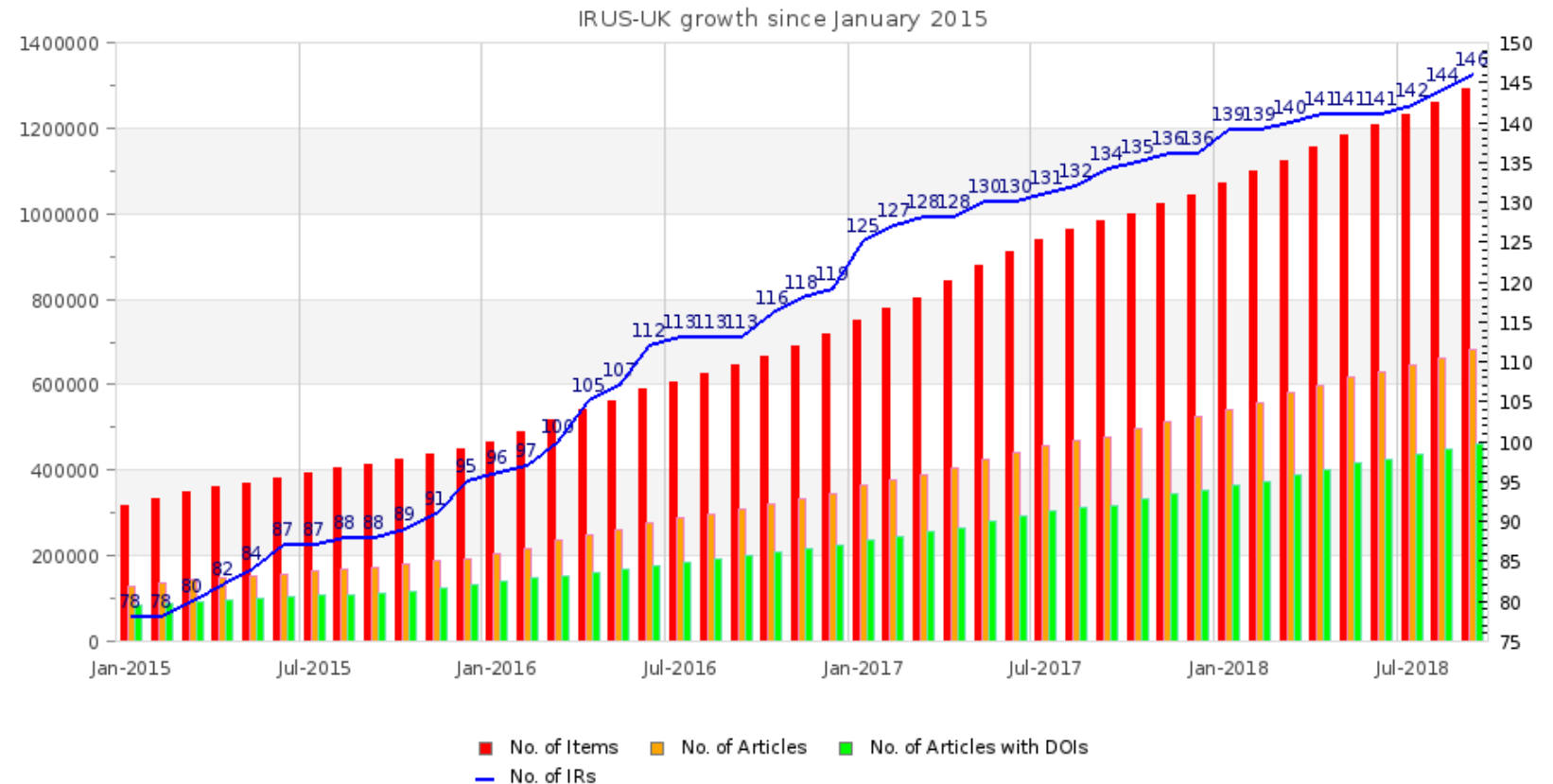
**IRUS-UK**

24/10/2018

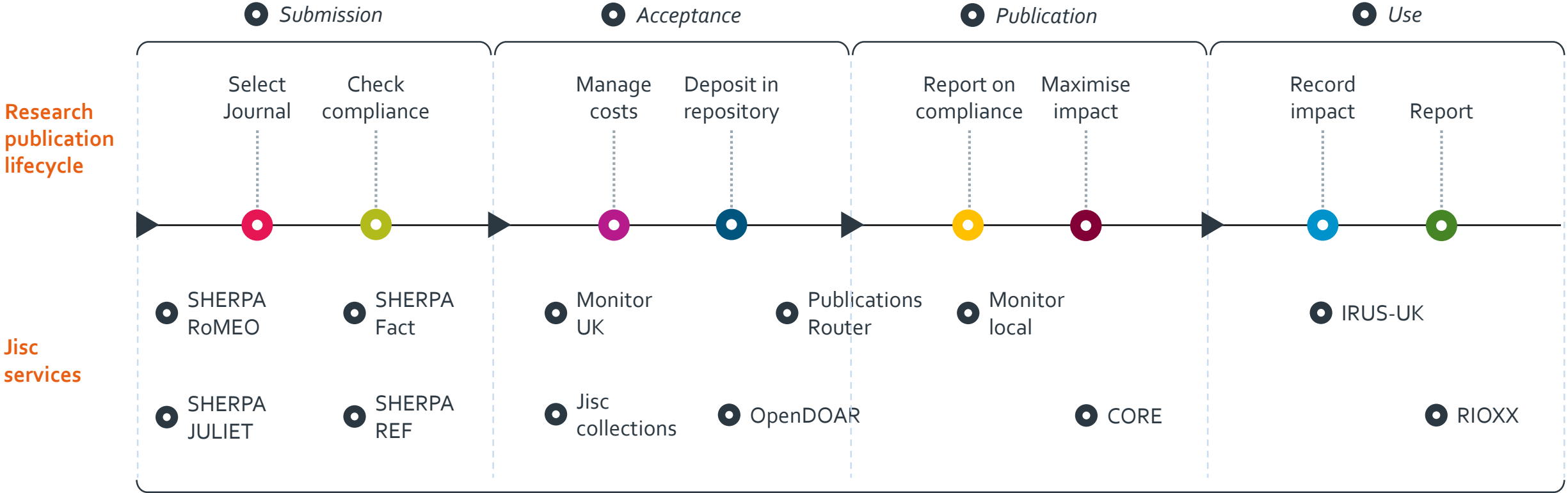
# IRUS-UK: the technical aspects

Paul Needham and Hilary Jones

- » IRUS-UK is a national aggregation service, enabling UK IRs to share and expose usage statistics at the individual item level, based on a global standard – COUNTER
- » Collects raw download data from UK IRs for all item types
- » Processes those raw data into COUNTER-conformant statistics
- » We have a growing number of participating repositories – currently 146



## Jisc Research Outputs Management Services



● Guidance, consultancy, technical support, and OA good practice



- » Files are being downloaded from repositories
- » How can IRUS-UK get information about those downloads from repositories?
  - › How do we get the usage data from IRs to IRUS-UK?
- » Then what do we do with that information?
  - › How do we process the raw usage data and convert to COUNTER-conformant statistics?
  - › How do we display, share, re-expose those statistics?

- » Data are pushed from repositories to IRUS-UK
- » We receive notification of download events
  - › On-the-fly as and when they occur
  - › In batches the day after
- » It's a relatively simple process:
  - › A file is downloaded from a repository
  - › Ping – a notification is sent to IRUS-UK server
  - › We handle all the processing from there
  - › We're responsible for the COUNTER-conformance side of things
- » It scales well
  - › More IRs = bigger logs, but no other real issues
- » Accomplished using the Tracker Protocol

- » Devised in collaboration with COUNTER, the specification is brief and straightforward
  - › <http://irus.mimas.ac.uk/documents/TrackerProtocol-V3-2017-03-22.pdf>
- » When a user clicks on a link to (i.e. downloads) a file from a Repository with the tracker protocol in operation, an OpenURL log entry is sent to a remote server for further processing.
- » The OpenURL log entry is based on a subset of the NISO OpenURL 1.0 standard *KEV ContextObject Format*.
- » The OpenURL string values
  - › must be URL encoded, with key-value pairs separated by &
  - › Appended to a base URL as a querystring



Element	OpenURL Key	OpenURL Value (example)	Notes
OpenURL version	url_ver	Z39.88-2004	Identifies data as OpenURL 1.0. String constant: Z39.88-2004 (Mandatory)
Usage event timestamp	url_tim	2010-10-17T03%3A04%3A42Z	Date/time of usage event (Mandatory)
Client IP address	req_id	urn:ip:138.250.13.161	IP Address of the client requesting the article (Mandatory)
UserAgent	req_dat	Mozilla%2F4.0+%28compatible%3B+MSIE+7.0%3B+Windows+NT+5.1%3B+Trident%2F4.0%3B+GoogleT5%3B+.NET+CLR+1.0.3705%3B+.NET+CLR+1.1.4322%3B+Media+Center+PC+4.0%3B+IEMB3%3B+InfoPath.1%3B+.NET+CLR+2.0.50727%3B+IEMB3%29	The UserAgent is used to identify and eliminate, by applying COUNTER rules, accesses by robots/spiders (Mandatory)
Item OAI identifier	rft.artnum	oai:dspace.lib.cranfield.ac.uk:1826/936	(Mandatory)
FileURL	svc_dat	https://dspace.lib.cranfield.ac.uk/bitstream/1826/936/4/Artificial_compressibility_Pt2-2005.pdf	(Mandatory)
HTTP referer	rfr_dat	https%3A%2F%2Fscholar.google.com%2F	The HTTP header field that identifies the address of the webpage (i.e. the URI) that linked to the resource being requested, which is used to help identify and eliminate accesses by robots/spiders. (Mandatory)
Source repository	rfr_id	dspace.lib.cranfield.ac.uk	(Mandatory)

- » A note about robots
- » COUNTER aims to count only 'genuine', user-driven usage
- » So, information about known robot downloads does not need be transmitted to the remote server
- » Before attempting to transmit the OpenURL, a check may be carried out to eliminate robots as defined in the COUNTER user agent exclusion list
  - › <https://github.com/atmire/COUNTER-Robots>
- » Reduces volume of data transmitted
- » But not essential as we also apply this filtering at our end

- » There are a number of software platforms offering Institutional Repository functionality, including:
  - › DSpace, Eprints, Fedora, Figshare, Haplo, Pure Portal, Worktribe
  - › and the list goes on . . .
- » Each of these platforms works differently under the hood meaning a single solution for deploying tracker functionality is not possible
- » We focused, initially, on DSpace and Eprints, which account for about two-thirds of UK repositories
- » We also now have Tracker implementations for:
  - › Fedora (Samvera), Haplo, Pure, Worktribe + a couple of bespoke IR systems

- » DSpace patches
- » We have Tracker patches, developed by @mire, for the three most recent versions of DSpace
- » To be able to install a patch, you will need the following prerequisites:
  - › A running DSpace 4.x, 5.x or 6.x instance
  - › Maven 3.2.2 or greater
  - › Git installed on the machine. The patch will be applied using several git commands
- » The patches are available at <https://github.com/atmire/IRUS>

- » Eprints plug-in
- » There is an Eprints Tracker plug-in for Eprints 3.2.x and 3.3.x
  - › developed by Eprints Services
  - › Additional contributions from John Salter (White Rose) and Alan Stiles (OU)
- » The plug-in is available from the Eprints Bazaar
  - › [PIRUS/IRUS-UK PUSH Implementation](#)
  - › A one click install (most of the time!)

## » Other softwares:

### » Fedora

- › (very!) flexible, modular, open source repository platform
- › Ruby gem for Samvera – contact us for more info if interested

### » Haplo, Pure Portal, Worktribe

- › Liaised with these vendors
- › Functionality implemented by their developers

### » Other platforms

- › We would welcome dialogue with interested vendors & developers
- › Contact us!

- » Tracker data is transmitted to a number of designated endpoints (URLs) on our server
- » Most implementations send messages on the fly
- » Pure sends a series of messages, the day after usage has occurred
- » Worktribe posts a single file containing a batch of messages, the day after usage has occurred
- » Usage data gets stored in daily log files associated with each endpoint

- » We run ingests on the daily log files, typically, 3 times a week
- » Ingest is a multi-script process, which adheres to the processing rules specified in:
  - › Release 4 of the COUNTER Code of Practice for e-Resources
  - › Release 1 of the COUNTER Code of Practice for Articles
- » When we first started just consisted of 3 scripts
- » Over time, as we've grown, encountered new situations, and enhanced functionality, it's become about a dozen scripts
- » Nearly time for a diagram . . . but first we need to discuss robots and rogue usage

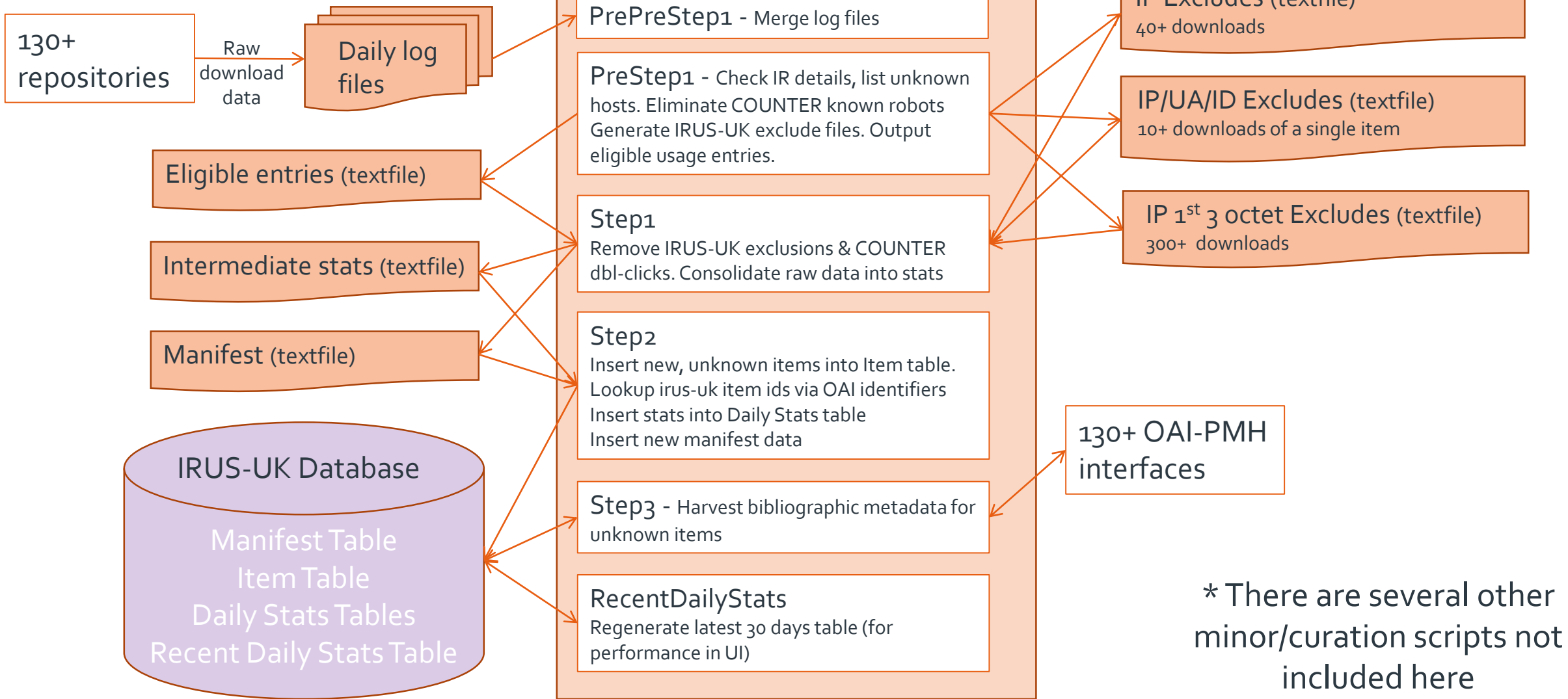


- » We remove robot entries using the COUNTER user agent exclusion list (<https://github.com/atmire/COUNTER-Robots>)
- » That's worked well in eliminating robot usage on traditional scholarly publisher websites behind an authentication/authorisation barrier
- » But it's not enough in the open access repository world
- » Besides 'good' bots like Googlebot, there are
  - 'bad' bots that don't declare themselves as bots but are mostly harmless
  - and a host of others: hackers, spammers, dictionary attackers, etc.
- » For those bad'uns we need additional filtering mechanisms above and beyond the COUNTER exclusions

- » So we additionally exclude entries from:
  - › known IP ranges of bots like Baidu Spider that don't declare themselves nicely
  - › Usage events with fake Google referrers
  - › IPs with 40 or more downloads in a single day
  - › IP/UAs with 10 or more downloads of a single item in a single day
  - › IP ranges grouped by the 1<sup>st</sup> three octets that have 300 or more downloads in a day
  - › These thresholds have been derived empirically by analysing our (extensive) logs
  - › During an audit review, the COUNTER auditors agreed that these appeared to be reasonable extra measures to remove robotic/rogue activity from our stats.

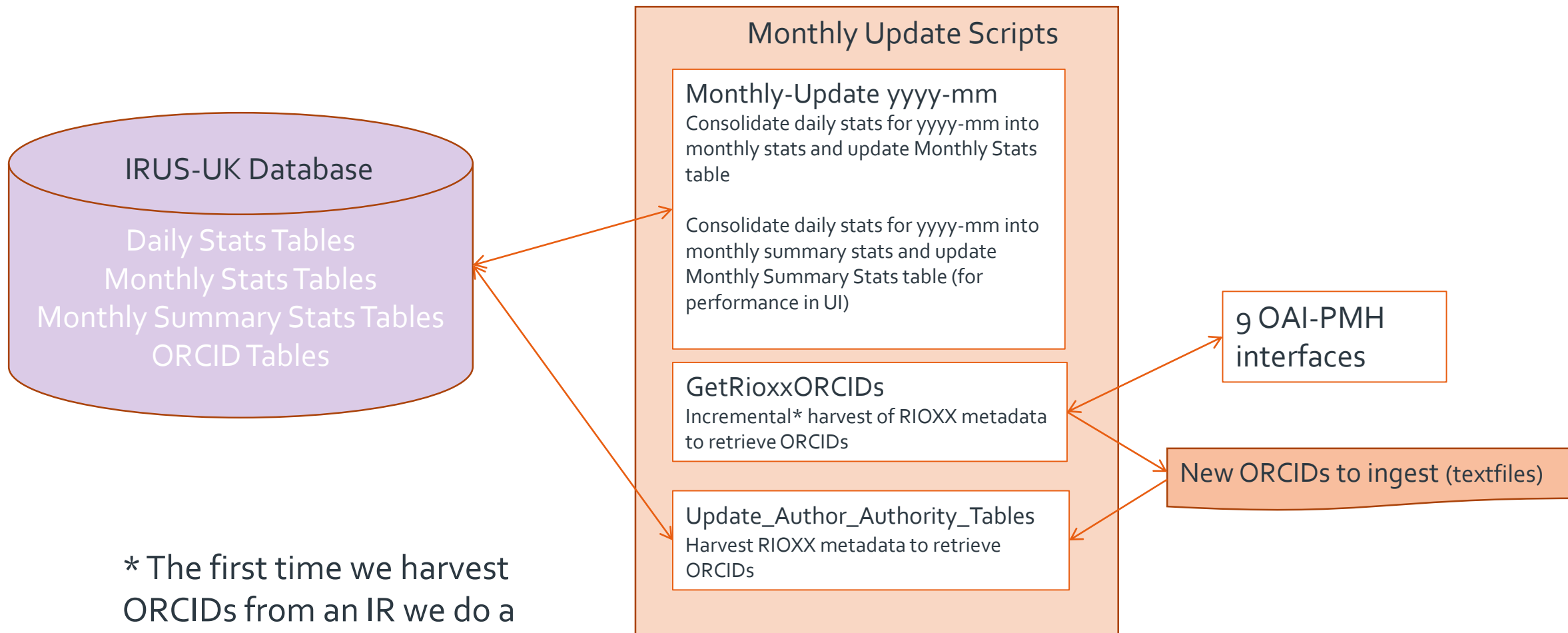
- » Scholarly publishers also implement their own measures, above and beyond the basic COUNTER exclusions, in their screening of usage data
- » It would be better, though, if there were a common, standard set of rules we could all reliably apply
- » To this end, we helped to form and are represented on the COUNTER Robots Working Group (COUNTER-RWG)
  - › The aim of the group is to devise 'adaptive filtering systems'
  - › sets of algorithms that will allow publishers/repositories/services to dynamically identify and filter out unusual usage/robot activity based on observed behaviours
  - › following a common set of rules, which will be incorporated into a future COUNTER Code of Practice.
  - › It has a wide-ranging membership, including representatives from publishers (Elsevier, Wiley), aggregators (EBSCO), repository software developers (DSpace, Eprints) and others
  - › Its work is ongoing . . .

# Daily Ingest



\* There are several other minor/curation scripts not included here

# Monthly Update



\* The first time we harvest ORCIDs from an IR we do a full harvest

## » Web User Interface -The IRUS-UK portal

- › Access currently behind Shibboleth authentication/authorisation
- › Wide range of views – slicing and dicing stats from the IRUS-UK database
- › Reports available for download as CSV/Excel spreadsheet files
- › Altmetric donuts and Dimensions badges for individual items 😊
- › Tableau Visualisations

## » SUSHI Lite API

- › Based on developments undertaken by the NISO SUSHI Lite Technical Report Working Group ([http://www.niso.org/workrooms/sushi/sushi\\_lite/](http://www.niso.org/workrooms/sushi/sushi_lite/))
- › RESTful: uses standard HTTP GET returning JSON
- › Allows machine to machine exchange of statistics
  - Full COUNTER or COUNTER-like reports
  - stats snippets which can be embedded into IR (and other) web pages
  - Used by OpenAIRE and others

## » Widgets – under development! Watch this space . . .

- » Our current service is based on release 4 of the COUNTER Code of Practice
- » A new R5 COP comes into force next year
  - › R5 focuses on improving the clarity, consistency, and comparability of usage reporting
  - › R5 reduces the overall number of reports, replacing many of the special-purpose reports that are seldom used with a small number of flexible generic reports
  - › R5 adopts the latest SUSHI format, JSON (instead of SOAP/XML)
  - › <https://www.projectcounter.org/code-of-practice-five-sections/abstract/>
- » Over the next 6 months, we'll be carrying out development work to transition IRUS-UK to R5
  - › Looking at amendments to the Tracker Protocol so we can count metadata views as well as file downloads
  - › Rewriting the ingest scripts to adhere to R5 processing rules
  - › Redeveloping the UI
  - › Adding COUNTER SUSHI API support



**IRUS team**

Email central helpdesk:  
[help@jisc.ac.uk](mailto:help@jisc.ac.uk)

Mention IRUS in subject line

Twitter: @IRUSNEWS

[jisc.ac.uk](http://jisc.ac.uk)



**Any questions?**