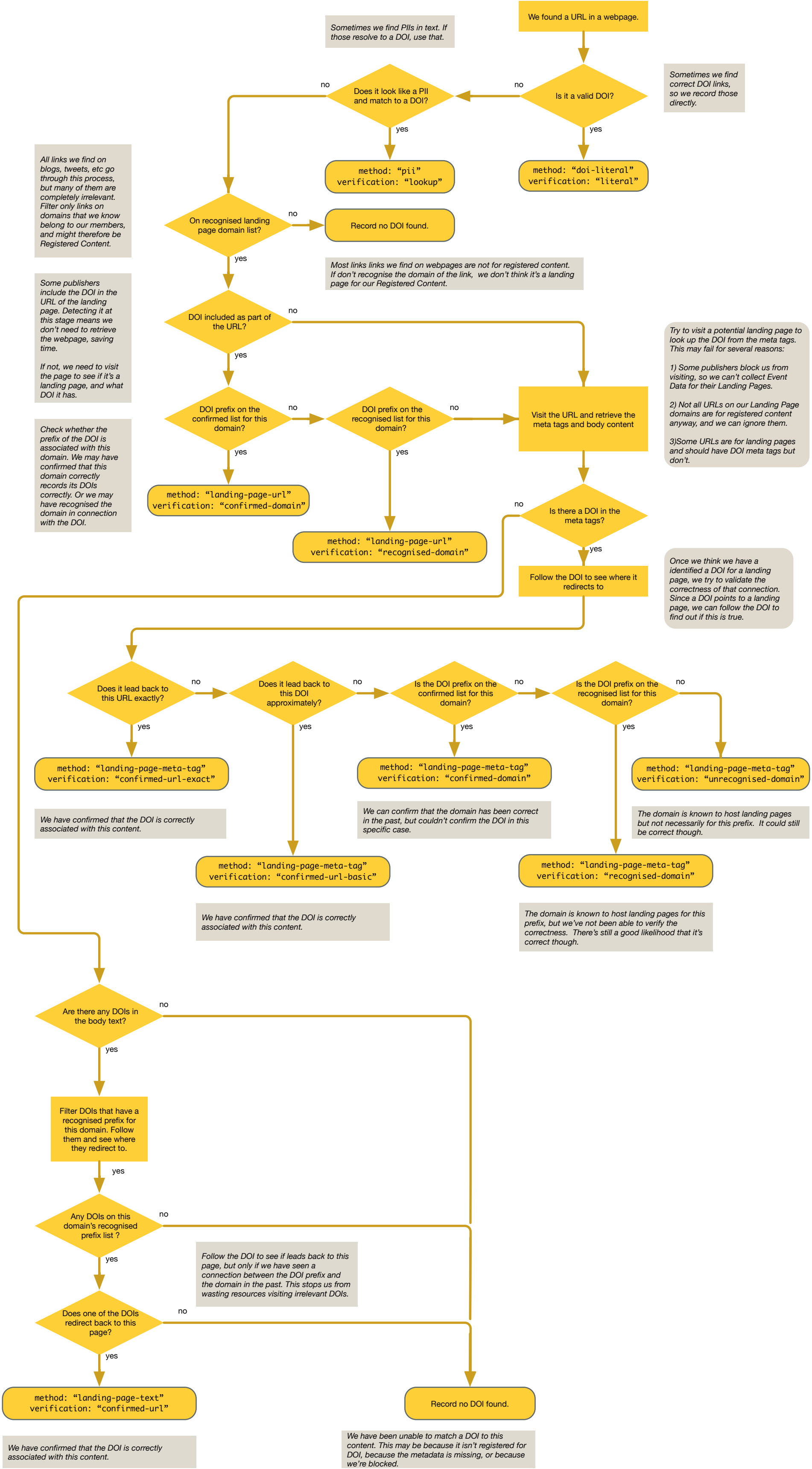
We found a URL in a webpage.
Sometimes we find PIIs in text. If those resolve to a DOI, use that.
Is it a valid DOI?
no
yes
Sometimes we find correct DOI links, so we record those directly.
Does it look like a PII and match to a DOI?
no
yes
method: "pii"
verification: "lookup"
method: "doi-literal"
verification: "literal"
All links we find on blogs, tweets, etc go through this process, but many of them are completely irrelevant. Filter only links on domains that we know belong to our members, and might therefore be Registered Content.
On recognised landing page domain list?
no
Record no DOI found.
yes
Most links links we find on webpages are not for registered content. If don't recognise the domain of the link, we don't think it's a landing page for our Registered Content.
Some publishers include the DOI in the URL of the landing page. Detecting it at this stage means we don't need to retrieve the webpage, saving time.
If not, we need to visit the page to see if it's a landing page, and what DOI it has.
DOI included as part of the URL?
no
yes
Try to visit a potential landing page to look up the DOI from the meta tags. This may fail for several reasons:
1) Some publishers block us from visiting, so we can't collect Event Data for their Landing Pages.
2) Not all URLs on our Landing Page domains are for registered content anyway, and we can ignore them.
3)Some URLs are for landing pages and should have DOI meta tags but don't.
DOI prefix on the confirmed list for this domain?
no
DOI prefix on the recognised list for this domain?
no
Visit the URL and retrieve the meta tags and body content
yes
yes
Check whether the prefix of the DOI is associated with this domain. We may have confirmed that this domain correctly records its DOIs correctly. Or we may have recognised the domain in connection with the DOI.
method: "landing-page-url"
verification: "confirmed-domain"
method: "landing-page-url"
verification: "recognised-domain"
Is there a DOI in the meta tags?
no
yes
Follow the DOI to see where it redirects to
Once we think we have a identified a DOI for a landing page, we try to validate the correctness of that connection. Since a DOI points to a landing page, we can follow the DOI to find out if this is true.
Does it lead back to this URL exactly?
no
Does it lead back to this DOI approximately?
no
Is the DOI prefix on the confirmed list for this domain?
no
Is the DOI prefix on the recognised list for this domain?
no
yes
yes
yes
yes
method: "landing-page-meta-tag"
verification: "confirmed-url-exact"
We have confirmed that the DOI is correctly associated with this content.
method: "landing-page-meta-tag"
verification: "confirmed-domain"
We can confirm that the domain has been correct in the past, but couldn't confirm the DOI in this specific case.
method: "landing-page-meta-tag"
verification: "unrecognised-domain"
The domain is known to host landing pages but not necessarily for this prefix. It could still be correct though.
method: "landing-page-meta-tag"
verification: "confirmed-url-basic"
We have confirmed that the DOI is correctly associated with this content.
method: "landing-page-meta-tag"
verification: "recognised-domain"
The domain is known to host landing pages for this prefix, but we've not been able to verify the correctness. There's still a good likelihood that it's correct though.
Are there any DOIs in the body text?
no
yes
Filter DOIs that have a recognised prefix for this domain. Follow them and see where they redirect to.
yes
Any DOIs on this domain's recognised prefix list ?
no
yes
Follow the DOI to see if leads back to this page, but only if we have seen a connection between the DOI prefix and the domain in the past. This stops us from wasting resources visiting irrelevant DOIs.
Does one of the DOIs redirect back to this page?
no
yes
method: "landing-page-text"
verification: "confirmed-url"
We have confirmed that the DOI is correctly associated with this content.
Record no DOI found.
We have been unable to match a DOI to this content. This may be because it isn't registered for DOI, because the metadata is missing, or because we're blocked.