

# Aggregation System Spec

updated 10 Feb 2012

## Background: Current Code

If you log in to [www.inknife.com](http://www.inknife.com) you will see the Aggregation Feed List at [http://www.inknife.com/rss\\_feedlist](http://www.inknife.com/rss_feedlist) (the code that drives this is located at `svn\branches\linux\apps\inknife\includes\rss_feedlist.include`).

This shows a list of XML and RSS-based site sources that we are pulling in:

The screenshot shows the iMedia Revenue Aggregation Management System interface. The page title is "iMedia Revenue" and the user is logged in as "Administrator". The navigation menu includes "Dashboards", "External Content", "Administration", "Newsletters", "Reports", "Help", and "Logout".

The "News Feeds" section displays "Refresh RSS Feed Content (may take a few minutes)" and "RSS Aggregation List | Web Site Aggregation". It shows "Showing up to 99 rows. | Show next 100 | Show all".

Below the navigation is a "Hide Filters" section with a table of filter options:

Feed Source	Interval	Text Search	Status	Curators	Flag
	All (30 days)		-Choose-	chester@si9.com rich@imediarevenue.com	<input type="checkbox"/>

A "Filter the List" button is located below the filter table.

The main content area displays a table of news feeds with the following columns: "Flag", "Source", "Headline", "Story Date", "Status", and "By (Disposition Date)".

Flag	Source	Headline	Story Date	Status	By (Disposition Date)
<input type="checkbox"/>	CBS New York	Convicted Sex Offender Michael Mele Admits Killing Aspiring Brooklyn Dancer	01/23/2012 17:06	None	
<input type="checkbox"/>	site:wsj.com manhattan - Google News	Weather Journal: Rain Washes Away the Slush - Wall Street Journal (blog)	01/23/2012 17:02	None	
<input type="checkbox"/>	NYT > N.Y. / Region	Javits Center Plans Are Deflating for Architects	01/23/2012 16:57	None	
<input type="checkbox"/>	manhattan - Google News	Sex offender admits killing aspiring dancer from Texas whom he picked up at ... - Washington Post	01/23/2012 16:56	None	
<input type="checkbox"/>	CBS New York	Sources: Fireworks To Blame For Fatal Queens Explosion	01/23/2012 16:55	None	
<input type="checkbox"/>	site:wsj.com manhattan - Google News	Man admits killing Texas dancer he met at NYC club - Wall Street Journal	01/23/2012 16:53	None	
<input type="checkbox"/>	CBS New York	Victorious Giants Return Home As NFC Champions	01/23/2012 16:50	None	
<input type="checkbox"/>	Top Stories - NY1.com	NY1 At 20: 2005 Report Looks At Factors Surrounding "Black Sunday" Firefighter Deaths	01/23/2012 16:38	None	
<input type="checkbox"/>	site:wsj.com manhattan - Google News	Real Estate News: Economists See Ways to Aid Housing Market - Wall Street Journal (blog)	01/23/2012 16:37	None	
<input type="checkbox"/>	Twitter / NotifyNYC	NotifyNYC: N/B Hutchinson River Pkwy closed at Westchester Ave (BX) d/t car accident. Whitestone Bridge traffic detoured to Bruckner Expy.	01/23/2012 16:36	None	

Shane redesigned this UI (much improved, with the filters on the left). For reference, see the interaction design update that Shane developed from my earlier wireframes:

## Curation

We need the ability to add editor's curation comments to an entry, right in the form. We can use Custommemo1 in the s1\_feeds table for this. This way an editor can click the dropdown, select "Accept", enter in a brief comment, and save the entry as "curated aggregation." As a result of accepting the entry, it would appear in the "curated aggregation feed" widget on the consumer-facing news site (which we still need to build), along with the editor's comments. This feedlist interface thus becomes the quick-publisher for aggregation items.

## Administration

It looks as if the feedlist is currently being driven off of the s1\_feeds table. Chester will plug in the Feeds module from NewsOrbiter instead. This is the module where we can add/edit/delete feed sources. The Feeds module is where we put in the URL of the feed and some basic metadata.

The Feed System then aggregates content according to a heartbeat, storing it in an aggregation table that is not a module table (because most of the data is throw-away). This aggregation data is used to build a queue of articles. The difference between the newsroom-specific aggregation queue and NewsOrbiter is that we'll be deleting feed content in the newsroom accounts after about 30 days.

Eventually we'll add sites, not just feeds, but for the demo we can just focus on feeds, which includes blog feeds.

## **News Topics and Feeds**

This interface is missing News Topics, which are in NewsOrbiter. We need to re-think the whole classification issue for aggregation items. Let's start with basics:

1. News feeds are module items in StageOne. There is a Feeds module. Each Feed can be classified with topics, just like the news stories:
2. Feeds also have other metadata, like Feed Source (New York Times) and URL (the URL of the RSS feed). By reference, all aggregated articles within a feed can "inherit" the metadata and classifications of their parent feed.

Now, we always faced two issues with this notion of inheritance, which is why the new system is going to be "curated" aggregation (meaning a human editor is involved). First, some RSS feeds are too generic to be tagged at all, and second, we would need to actually look into an article and search for keywords, comparing against a master taxonomy, in order to apply tags programmatically to the aggregated items. We can only do this once the master taxonomy is created. So this means we need to tag aggregated articles by hand for a while anyway.

So the Feeds can supply some metadata, and the individual aggregated items have other metadata (like headline and date). An editor called a "curator" will fill in tags, but only when he sees items he wants to aggregate. (80% of the feed content will likely not be of interest.)

## **Aggregated Articles as Link VITALs**

This means that a percentage of aggregated articles need to be "accepted" and then classified. Specifically, they need to become Link vital, because in order to drive the customer-facing UI, we need the aggregated articles to have their own tags, entered by a curator.

The most important classification is News Topic, as this drives everything. (This is the same topic defined for articles in the newsroom - classifications like Sports, Finance, Politics, etc.). This is important for the consumer-facing interface, because for each item "accepted" they have to appear according to its Topic in the consumer-facing news aggregation widget.

As the main organizing feature, topics can do a few things: the organization may want to have each topic go with a separate web site. Or the widget may be pre-topic-filtered, meaning that the Sports page just shows the widget with only Sports content, while the Business page only shows the widget pre-filtered with business content. To be really sophisticated, the widget could appear on each story page, showing related content from outside sources based on the tags!

So each aggregation item marked "accepted" minimally needs to be classified for News Topic (inherited from the Feeds parent item, if the parent Feed is classified to a topic). We need to code the page such that each "accepted" aggregation item is copied into the Links module. I have a set of screens I use to do that manually already (curation.cfm). But we can do this programmatically pretty easily.

So the current aggregation queue would just be a list of potential link vitals, not the list of actual aggregated Links vitals. The aggregated content would appear in the Links module and could then be classified using standard StageOne mechanisms:

**TAGS**

News Topics:  
 News Topics (1 of 14 chosen) ▼  
 Business & Economy

Who?:

Where?:

What?:

Why?:

(BTW this gives us a nice rich purpose for Links, with three types of Link items: plain URLs, aggregated articles, and Flash interactives).

So the aggregation system would enable you to click "Accept" on an aggregation item, and in a div popup, we can show an Impress form for the Link vital, based on a code family story\_links (I am not against the idea of a custom impress\_aggregation code family that has the right fields for an aggregation item, but I don't yet think it's necessary). In other words we can completely leverage our existing infrastructure.

Here are some wireframes based on the old visual design, not Shane's cool new design:

Flag	Source	Headline	Story Date	Status	By (Disposition Date)
1	<input type="checkbox"/> CBS New York	Convicted Sex Offender Michael Mele Admits Killing Aspiring Brooklyn Dancer	01/23/2012 17:06	None	
2	<input type="checkbox"/> site:wsj.com manhattan - Google News	Weather J Headline: <input type="text"/>	01/23/2012 17:02	None	
3	<input type="checkbox"/> NYT > N.Y. / Region	Javits Cer Link (URL): <input type="text"/>	01/23/2012 16:57	None	
4	<input type="checkbox"/> manhattan - Google News	Sex offend 3rd Party Credit: <input type="text"/> Washingt	01/23/2012 16:56	None	
5	<input type="checkbox"/> CBS New York	Sources: I Description: <input type="text"/>	01/23/2012 16:55	None	
6	<input type="checkbox"/> site:wsj.com manhattan - Google News	Man admi	01/23/2012 16:53	None	
7	<input type="checkbox"/> CBS New York	Victorious	01/23/2012 16:50	None	
8	<input type="checkbox"/> Top Stories - NY1.com	NY1 At 20 Deaths Why?: <input type="text"/> As NY1 N Meminger &quot;Bla two separate fires.	01/23/2012 16:38	None	
	<input type="checkbox"/> site:wsj.com manhattan - Google	Real Estate News: Economists See Wave to Aid Housing Market - Wall Street Journal	01/23/2012	None	

Note that we can't possibly make all of Shane's design work just yet, plus it needs some re-thinking in a few places, especially as he conceived of it as a customer-facing browsing tool, which it isn't. It is an editorial curation tool that helps you scan a hundred headlines at a

time, several times per day, deleting and bookmarking and aggregating as needed. But here are the basic ideas:

1. Click the Aggregate button to start. We probably don't need the old dropdown I coded; I like the delete (✖) and bookmark (☆) and already aggregated (✓) icons, but if you need to save time, feel free to keep the dropdown since it is already Ajaxed (you'll need to add the value of "aggregated").
2. Whatever the gesture, it has to be easy to scan the page through dozens of headlines and see what has already been aggregated and what might still be of interest.
3. When you click to aggregate, a div opens up with the Link vital form as a fly-out. It is pre-populated with the content from the aggregation item (headline, description, date, etc.).
4. The Link vital (we would use a code family impress-aggregation) would include the field for pointing to the thumbnail image and other things specific to an aggregation item. Or we could just use the story\_links code family if we think that it makes sense to always have tags on a Link item.
5. Ideally the aggregation queue is updated only after the div is saved, but if that's a hassle, for now I'm cool with it being updated as soon as you click the select box, which is how it works today (ColdFusion Ajax). If they cancel the div, I suppose then they can just unmark the select box. If you're wondering why we need the status dropdown at all, it's because the queue needs to know what happened, as it's sort of a report.
6. Once saved, the Aggregated Link vital is available in the shuttlecraft Archive under Links. It doesn't have to have a story associated with it; it can be displayed using the consumer-facing widget instead.
7. If the story does have Link vital that is type=aggregation, then we just need to figure out how to display it in the story page. Not a big deal; maybe if it's just one we show it as a "related link" and if there are more than one we show them in the consumer-facing widget? (Or maybe they always display as related links, depending on what the customer wants).
8. Since the Story quick-form now needs to accommodate aggregated Link fields, we can have the code family story\_links cover all types of links, including aggregation. These would be:
  - a. Headline
  - b. Link (URL, Permalink)
  - c. 3rd Party Credit
  - d. Source
  - e. Description
  - f. Thumbnail URL
  - g. Curation Comments

- h. Published Date
  - i. News Topic and Tags (all of the tag classifications used by the news stories, with the div closed by default, so as not to confuse folks who just want to put in an interactive). The div would be called "Aggregation Tags".
9. Note that we don't need to track the parent Feed Source or anything like that in the link vital. Once an article is aggregated, the feed is irrelevant (and of course the news agency could change the feed anyway).

### **Single Page Aggregation**

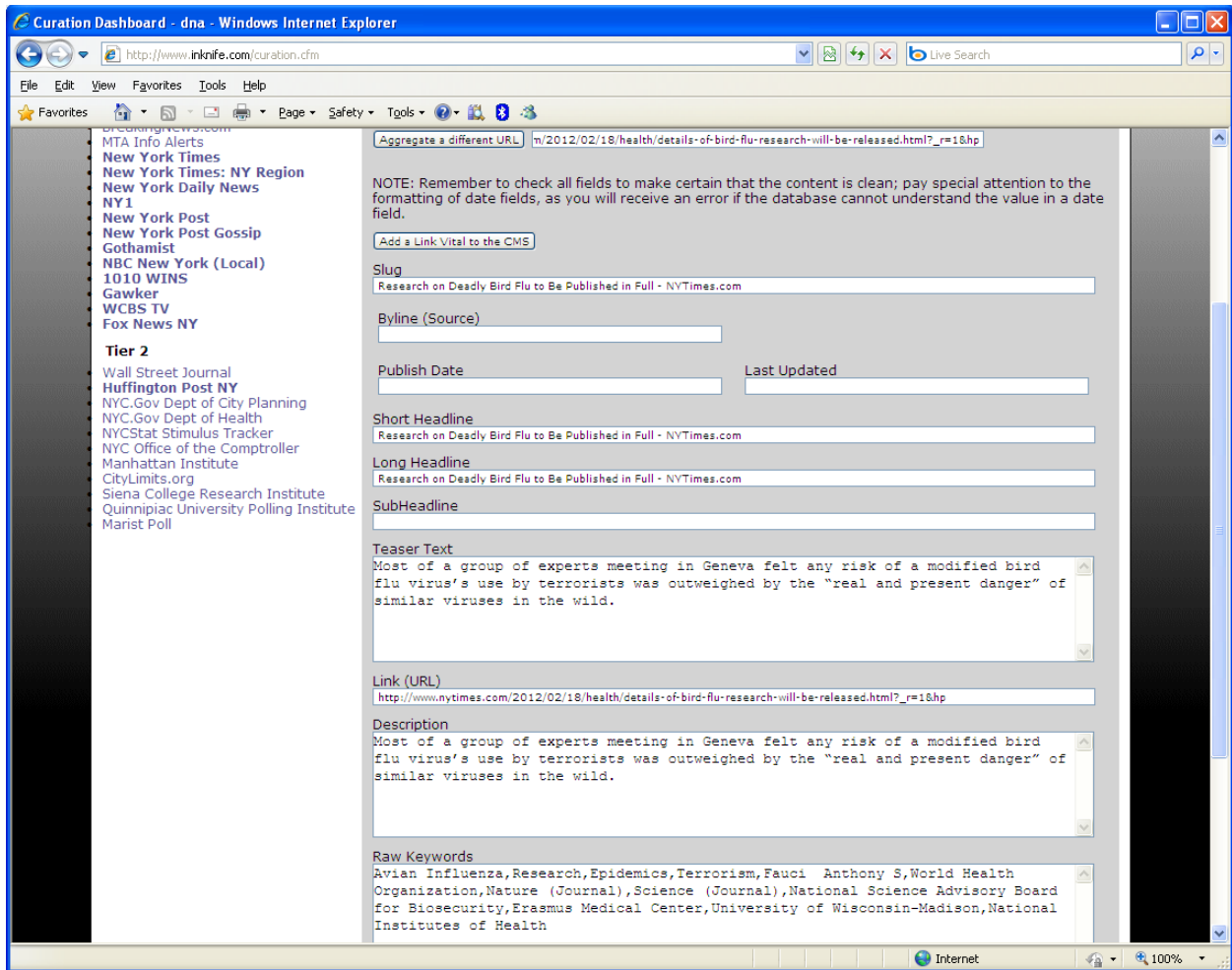
Based on the discussion above, we can think more clearly about the single-page aggregation. Here, we can do one of three very simple things:

1. Create a story and add a Link vital.
2. Create a shuttlecraft page for the Links module and enable curators to add aggregation links as needed (outside of stories). They would just copy and paste the headline and description by hand and add any curation comments, specify the tags, and presto they're done.
3. Go to the Web Page Aggregation utility I built, and snarf the page, then save it as an aggregation item, again setting Topics and Tags manually. This system is still working I believe, although curation.cfm needs some serious layout work.

(Someday we'll be able to parse the tags automatically based on a lookup against the controlled vocabulary, which consists of the existing Tags.)

On rss\_feedlist.cfm there is a subnav link to Web Site Aggregation which manages any single (arbitrary) web page you want to aggregate. This is an important part of the aggregation system. Right now it is being driven off of an EWP page, <http://www.inkknife.com/curation.cfm> (code is located in svn\branches\linux\apps\inkknife\includes\curation.include). What we need is to clean up this interface a bit and to use it as a feeder to the Links module.

This page has a button to "Add a Link Vital to the CMS". This form already enables the editor to enter in their curation comments, but I don't think this make sense; it should just post to the Link vital where they can do their curation.



## Consumer-side Aggregation Widget

We need a widget that we can plug into any web site. It would not be the full NewsOrbiter interface. It needs to show a list of curated article headlines. When you click the headline, a fly-out shows you the details:

- a. Headline as hyperlink to article ( we can add a Link icon)
- b. Byline Credit
- c. Source
- d. Description
- e. Thumbnail
- f. Curation Comments
- g. Published Date

The widget will potentially need to support three states:

1. Site-wide (meaning it needs to display topics as a filter or as tabs)
2. Topic-specific (meaning that it sits on a topic/category page or a story)
3. Story-specific (the same as topic-specific for now, but in the future it could list articles hand-selected for a given story)

In all cases, the fly-out described above would be the same.

A site-wide version would be handy to demo on the home page. This would be sorted by topics, probably as tabs, although widgets aren't usually that wide so we need to think about this. Or we could make the topics into filters and just have a scrolling list of article headlines.

Regardless, click on a headline and you'd see the aggregated content: the source and description and date and editor's curation comments, etc.

The Topic-specific version would be the most useful, and is what we'll want to deliver to Newbay, so we need to build this one first. It's the simpler version: just a way to show headlines, with a Topic as its title:

### **Guitar News**

Fender releases retro Mustang  
Korg delivers new sound gizmo  
Marshall blasts its way to the top  
Keith Richards New Book

or better, with a news source or logo:

### **Financial News**

Financial Unrest in Greece (Bloomberg)  
Irish Markets Rebound (Irish Times)  
iMedia Saves the World (Guardian)

---

MY NOTES ON SHANE'S NEW AGGREGATION UI; FROM EARLIER EMAIL, BUT MOST ARE STILL RELEVANT

Shane, this is very good.

I had not imagined that NewsOrbiter would have the same UI as the curation aggregation tool, since one is consumer facing and the other is a tool for editorial. But I think you did a good job merging most of the functionality. Still, there would be minor differences between the two systems. Here are my first-reaction notes:

- It's a little hard to scan the headlines; they need to pop out a bit more.
- NewsOrbiter (the consumer facing UI) doesn't have "reject" and "accept" attributes ("star" is like "bookmark" so that one works in both systems—the only curator in NewsOrbiter today is the user, but Chester and I do have a model for "curated content" so this will make sense once we add that part).
- I think we need a control that gives options for how many results to display at a time. 10 seems too little, with too much pagination. For consumers 20 to 100 is reasonable, for editorial 100 at a time is the likely minimum, because they are not casual users (it's their job to review massive amounts of aggregated content)
- Rejected stories should probably not be shown in the UI (you reject them to clear out the clutter)
- The list of curators wasn't in NewsOrbiter, but like the editorial site, the consumer facing site can show stories flagged (and perhaps annotated) by editors
- The "aggregate" option doesn't exist in the consumer-facing UI.



- Is the headline an obvious enough open/close toggle? Maybe we need this to be icon-driven.
- The “Web Site Aggregation” (+) icon doesn’t really make sense since it’s a link to an editorial administrative tool (it isn’t in NewsOrbiter)
- The top right navigation is really cool, but the consumer-facing tool is really a stand-alone product, and the editorial version is a tool within the overall menu system, so shouldn’t it use that nav? (Or maybe we should incorporate these icons into Impress? Or maybe we should keep this as a “utility” that doesn’t share the same nav? Have to think about this.)

Regards,  
Rich

**From:** Shane Carmody [<mailto:shane@imediarevenue.com>]  
**Sent:** Wednesday, January 11, 2012 8:44 AM  
**To:** Rich Julius  
**Cc:** Stephen Collins; Derek Burke  
**Subject:** News Agg. Mockups

Hi Rich,

I'm sending you the mockups I did for the News Aggregation Tool, can you review them? I'm attaching a zip file which contains 10 images and 1 PDF which explains each image. Images 01 and 09 are clean images without any annotations.

I've combined the NewsOrbiter website with the News Aggregation Tool at [http://www.inkknife.com/rss\\_feedlist.cfm](http://www.inkknife.com/rss_feedlist.cfm). I've cleaned up the UI and made it much more compact.