This section of the presentation on IOTA will focus on the Completeness Index, its validation and optimization.
We will start with an overview of the premise behind the completeness scores and completeness index and the corresponding element weights. Next we will review how we went about validating the concept, followed by the technique used to optimize the element weights. And finally we will look at the results of this work, our conclusions and the next steps for IOTA.
The premise behind IOTA

- **Completeness Score** is the measure of the “completeness” of a single OpenURL.
- **Completeness Index** is attributed to the content provider as an overall measure of the completeness of their OpenURLs.

Let's start with a couple definitions...

The Completeness Score is the measure of the completeness of a single OpenURL. And by completeness we mean the number of metadata elements provided in the OpenURL out of a desired or core number.

The Completeness Index is a number that is attributed to a content provider (OpenURL “referrer”) to measure the completeness of their OpenURLs in aggregate. It is essentially an average of the completeness scores of OpenURLs coming from that content provider.
The premise behind IOTA

- The Completeness Score is calculated by “weighing” the elements provided in the OpenURL based on their importance in target links
- Some elements are more important than others and will have a higher weight
- Completeness Score equals the sum of weights of elements found divided by the maximum score possible

The Completeness Score is a number between 0 and 1 that measures the “completeness” of the elements in the OpenURL, with 0 indicating no elements provided and 1 indicating all desired elements were there – a perfect OpenURL.

It is a weighted score with some elements carrying more weight than others – the element’s weight is based on its importance in target links.

<CLICK>
When you look at the syntaxes for target links, you will see some elements like ISSN, Volume and Issue appearing more often than elements like article title or author. Common sense tells us that, if an OpenURL is missing an element that is needed in many Target Links its probability of failure is much greater than if it was missing an element needed by only a few targets.

<CLICK>
So each element is assigned a value based on its “importance”...we call this the element weight. Calculating the Completeness Score is done by taking a sum of the “weights” for each core element found in the OpenURL and dividing that sum by the maximum possible score. Basically comparing the data in a OpenURL to a “perfect OpenURL”
The premise behind IOTA

- Simple example assuming equal element weights

<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
<th>Weight</th>
<th>This OpenURL</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATitle</td>
<td>Article title</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>AuLast</td>
<td>Author's last name</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Date</td>
<td>Date of publication</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>ISSN</td>
<td>ISSN</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Issue</td>
<td>Issue number</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>SPage</td>
<td>Start page</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>Journal Title</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Volume</td>
<td>Volume number</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td></td>
<td>8</td>
</tr>
</tbody>
</table>

Let's run through a quick example. This table shows the core elements for an article link... and for the simplicity of this example we will assume all elements are equally important so each gets a weight of 1 – a perfect OpenURL will get the maximum score of 8.
The premise behind...

- Simple example assuming complete

Completeness Score...
(Total for This OpenURL)
Total Weights
5 / 8
= .625

Now let's look at some OpenURL elements... In this OpenURL we have...

<CLICK>
Date ... so we add one point
<CLICK>
ISSN... add another point
<CLICK>
Volume... another point
<CLICK>
ISSUE... another
<CLICK>
And Article Title... and another point
<CLICK>
... the result is a total of 5 points.
<CLICK>
The calculation is Sum of the weights for this OpenURL divided by the total for all weights
<CLICK>
Which is five divided by 8
<CLICK>
Or .625
Determining the weights

- Initial approach
  - Frequency of element occurrence in target link templates
  - Combined with reasoning

Since not all elements are equally important, giving each a weight of 1 is not reasonable. So how did we come up with the initial set of weights? We looked at the link syntaxes for over 300 target link and did an occurrence count of each of the core element and used that to influence the weights. Then tweaked based on what looked about right.
Here is what we came up with...

<table>
<thead>
<tr>
<th>OpenURL data element</th>
<th>Description</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATitle</td>
<td>Article title</td>
<td>1</td>
</tr>
<tr>
<td>AuLast</td>
<td>Author's last name</td>
<td>1</td>
</tr>
<tr>
<td>Date</td>
<td>Date of publication</td>
<td>5</td>
</tr>
<tr>
<td>eISSN</td>
<td>Online ISSN</td>
<td>3</td>
</tr>
<tr>
<td>ISSN</td>
<td>Print ISSN</td>
<td>3</td>
</tr>
<tr>
<td>Issue</td>
<td>Issue number</td>
<td>3</td>
</tr>
<tr>
<td>Jtitle</td>
<td>Journal Title</td>
<td>1</td>
</tr>
<tr>
<td>Pmid</td>
<td>PubMed ID</td>
<td>8</td>
</tr>
<tr>
<td>SPage</td>
<td>Start page</td>
<td>3</td>
</tr>
<tr>
<td>Title</td>
<td>Journal Title</td>
<td>1</td>
</tr>
<tr>
<td>Volume</td>
<td>Volume number</td>
<td>3</td>
</tr>
<tr>
<td>DOI</td>
<td>Digital Object Identifier</td>
<td>8</td>
</tr>
</tbody>
</table>
Individual elements got weights from 1 to 5 with identifiers getting a higher score of 8.
A couple of comments... Most link resolvers will handle either or both print and online ISSN — if you have one, the resolver will look up both -- so score each separately isn’t a reflection of reality.
Also, identifiers, like PubMed ID and DOI are used by most link resolvers to look-up the full citation in PubMed or CrossRef—thus if you have one of these identifiers in an OpenURL you really have access to all the core data elements.

Giving an OpenURL with just a DOI fewer points than an OpenURL with DOI and other elements isn’t really valid since at the end of the day, the link resolver data enhancement will make them equal (in fact the DOI-only one might work better if there were any conflicts in metadata values.)
Validating the Completeness Score

- Use real OpenURLs and a commercial link resolver. (*tested with LinkSource and 360-Link*)
  - Remove institutional holdings as a limit to resolution
- Process each OpenURL through the link resolver to determine “Success”
  - Score one point for finding at least one full text target
- Calculate the completeness score for each OpenURL
- Look for a statistical correlation between the completeness score and the success score

But that is what we started with and we needed to see if we could statistically justify these numbers. So we ran some tests...

Aron Wolf from Serials Solution and I each ran a series of tests using the same OpenURLs on our respective link resolvers.

The first thing we did was eliminate library holdings as a variable by using the entire knowledge base as out “collection”. Remember we were testing whether there is enough of the right data elements to find full text, not if the library had a subscription.

<CLICK>
We then took a sample of some 15,000 OpenURLs — picked at random. We ran each through our link resolvers to see if the OpenURL would populate a full text target. We gave the OpenURL a “success score” of 1 for finding a full text target and “0” for not.

<CLICK>
The theory is that if the completeness of the OpenURL affects its success, which means there should be a relationship between “success” and the “completeness score”… so we also calculated the “completeness score” for each OpenURL.
<CLICK>
Then we looked for a statistical correlation between the two.
After running the numbers this is what came out of the LinkSource test... The graph shows average Completeness Score and average Success score by “referrer” — (OpenURL source). You can see some tracking but not overwhelming.

When we calculated the Correlation Coefficient between these two scores across all 15,000 OpenURLs in the test, the result was .43. Which indicates some level or relationship but, again, not particularly strong.
A Statistical Approach to Determining Element Weights

- Select a set of “perfect” OpenURLs
  - include all key data elements and resolve to full text
- Perform step-wise regression
  - Test failure rates for each element by removing that element
- Use failure rates as basis for weights
- Use new weights to test for correlation between weights and success for larger sample

We needed a better way of determining the element weights, so we sought help from Phil Davis – a researcher with some experience in statistical modeling. Phil’s suggestion was to perform stepwise regression to see the effect of individual elements on a sample of OpenURLs. And that is what we did...

We started with a set of “perfect” OpenURLs – ones that not only included all core data elements, but that also resolved to match a full text target on both LinkSource and 360 Link... we used a set of 1500.

<CLICK>
We then ran several series of tests where we ran the OpenURL past the link resolver with a different element removed for each test series.

<CLICK>
We recorded the success (or rather failure rates) associated with each element. The elements with the higher failure rates are more important to the success of the OpenURL than the ones with lower failure rates.

<CLICK>
We then used the failure rates as a basis for weights.

<CLICK>
Then we used the weights and re-ran our 15,000 sample test.
So how’d it turn out? Again, here are numbers for LinkSource.

You can see Volume was a key element with 74% of OpenURLs failing when it was removed.

Author last name was not very important with less than a 10th of a percent failure rate

Date was surprising low too. This could be for a few reasons — the level of forgiveness in the holdings matching logic (e.g. treat no date as “any date”), the ability for the link resolver to discover the date by looking up the article citation in the knowledge base using volume/issue/start page coupled with the fact that a lot of full text providers don’t use date explicitly in the outbound links.
We created article weights.

<Click>
Rather than use raw failure rates, we used logarithmic values of the failure rates – the number of failures per 10,000.
Then we ran our 15,000 record sample again. You can see from the graph that average completeness score and average success score for the OpenURL providers align very closely, and the Correlation Coefficient of these two values across all 15,000 test OpenURLs is .80 – which indicates a strong correlation. Good news for the test.

This tells us that the Completeness Index can be used as a predictor of OpenURL success from a particular content provider – a low Completeness Index is a good indicator there is a problem.
Notes

Testing the same OpenURLs on 360-Link results in different numbers but consistent trends. Differences may be attributed to:

- Variations in metadata enhancement techniques
- Strictness in target link rules (e.g. required elements before link shows – tied to level of forgiveness of target)
- Link syntax used for target

As I mentioned, similar test were conducted by Aron Wolf using 360 Link. The same OpenURLs were used and Aron was able to validate the concept of the Completeness Score; however, the failure rates were not exactly the same as for the LinkSource test; therefore, Aron’s weights were not the same... So why the difference?

Some possible explanations would be the variations in metadata enhancement -- the ability to “correct” weaknesses in an incoming OpenURL.

Strictness in the target link rules. For example, some targets will display just the journal home page if insufficient metadata is available – in real life this is probably better than no link because the user can navigate to the full text from there. For the LinkSource test, for example, we had removed a lot of the forgiveness because we wanted “success” as getting directly to the full text.

And link syntaxes is another possible variable. For example, JSTOR has an OpenURL-based syntax and another based on SICI code. The data you need to each is different therefore depending on the syntax used, the weights for corresponding elements would be different.
96.3 percent of the OpenURLs in the LinkSource test were successful in populating either a full text target or an ILL form (with a sufficient data that the ILL department could complete the request). This seems pretty good, so why is OpenURL viewed as problematic?

<CLICK>
- Part of it is perception -- what constitutes success in the eyes of a link resolver may not constitute success in the eyes of an end user. Remember our tests were done where holdings were not a factor (as if our test library subscribed to everything in the world). If you add specific library holdings into the mix, then more of the links will go to ILL forms -- many end user consider a link to an ILL form a broken link -- this is the perception problem.

<CLICK>
- A couple thoughts to combat this... if your discovery provider systems (the source of the OpenURL) allow, use their holdings features to change the wording of the OpenURL link.
  - Use “Link to full text” if item is from a journal in the library’s online collection;
  - use, “check library collection”.
  - if it is the print collection; or, use “request from library” for everything else. This way the end user’s expectations are controlled and the perception of the technology improved.
Conclusions

- Step-wise regression approach to element weights works
- Completeness Index scores can be correlated to actual OpenURL “success”
- KB and resolver technology influence results and prevent a universal set of element weights

The Completeness Index is a mechanism individual link resolver vendors can use to provide metrics to help improve their service quality

Our conclusions...

The step-wise approach to determining element weights works. The completeness index scores do correlate to OpenURL success.

However, Knowledge Base contents and resolver technology influence the results and thus prevent a universal set of weights and scores.

<Click>

The Completeness Index is a mechanism individual link resolver vendors can use to provide metrics to help improve their service quality by identifying those OpenURL sources that are more problematic.
Other takeaways

Several factors involved in perceived “link failure”:
1. Bad or missing metadata in the OpenURL link
2. Inaccurate holdings data within the resolver’s knowledge base
3. Flexibility of syntax to the target
   - e.g., target supports at least two: OpenURL syntax, DOI link, proprietary link structure
4. Flexibility of resolution logic at the target
   - i.e., target finds way to create link using available data when some data missing or wrong
5. User expectations
   - e.g., link resolver provided link to OPAC or ILL form, but user was expecting full text

- IOTA focused on (1)
- KBART working on (2)
- Education of content providers could address (4)
- Displaying OpenURL button only if full text available could address (5)

Other take-aways...

(as indicated on the screen)
What’s next for IOTA

- Continue offering public access to reports on element frequency
- Publish technical report on work to date
- Publish recommended practice for calculation and use of completeness scores for link quality assessment by link resolver vendors
- Continue work as a NISO standing committee for at least one more year

THANK YOU!

And what’s next for IOTA is...

(review what is on the screen)