



SAMALANDER

Apple-to-Apple Metrics for DITA

Meaningful metrics for writers, managers, across
business units and businesses.

This talk is about

- Defining the metrics problem in business terms.
- Clearing up some terminology confusion.
- Defining a technology- and time-independent baseline for metrics.
- Look at some examples.



Who we are ...

- Small company focused on independent contractors, small companies and small groups in multinationals.
- For one customer we increased writer productivity from 1,350 pages/year to about 10,000 pages/year or 1,650,00 words over 4 years

= 7.5 x increase in writer productivity
(but see slide 32 ...)



SAMALANDER



METRICS IN BUSINESS TERMS

Apple-to-Apple Metrics for DITA: CIDM Content Management Strategies / DITA Conference, Rhode Island, 2013-04-15-17

The Problem ...

- 7.5X productivity increase is impressive but ...
 - The increased productivity
 - means nothing in budget discussions
 - means nothing in workload discussions.



The Problem ... TechDocs Budget

- Why do the metrics mean nothing?
- **Because there is no way to know if 7.5 times increase in productivity is real.**
 - What is a page ? Are pages comparable?
 - What is a topic? Are topics comparable?
 - Is DITA necessary? Why not use MS Word?



More Serious Problems are ... #1

- Questions about metrics
 - Are the metrics **STABLE** from one year to the next?
 - Are they useful for planning and budgeting?
 - What if the writer(s) don't agree with the metrics?
 - What if upper management doesn't know anything about documentation?



More Serious Problems are ... #2

- Does everyone involved understand?
 - If you explain enough, maybe, BUT
 - They may say they understand, but don't really, OR
 - They may forget from one evaluation or budget cycle to the next.



Most Metrics Suffer from DRIFT ... WHY?

- Page count: because of formatting changes, adding a level to the ToC ...
- Topic size: because of increasing reuse, rewriting for reuse ...
- Topic size: because of increasing granularity, rewriting for reuse ...
- Difficulty: because of improved relationships with SMEs ...



Most Metrics Suffer from DRIFT ... Result

- MOST (currently all?!) metrics suffer drift because they are based on **abstract concepts**:
 - topic,
 - chunk,
 - paragraph,
 - section,
 - procedure.



We need an easy-to-understand baseline metric that is ...

- **Stable over time,**
- **Stable over work habit changes,**
- **Stable over technology changes such migrating from disk-based authoring to a CMS then a CCMS**
- **Stable even as our position in the DITA maturity model**



SAMALANDER



CLARIFYING SOME SOME TERMINOLOGY CONFUSION

Apple-to-Apple Metrics for DITA: CIDM Content Management Strategies / DITA Conference,
Rhode Island, 2013-04-15-17

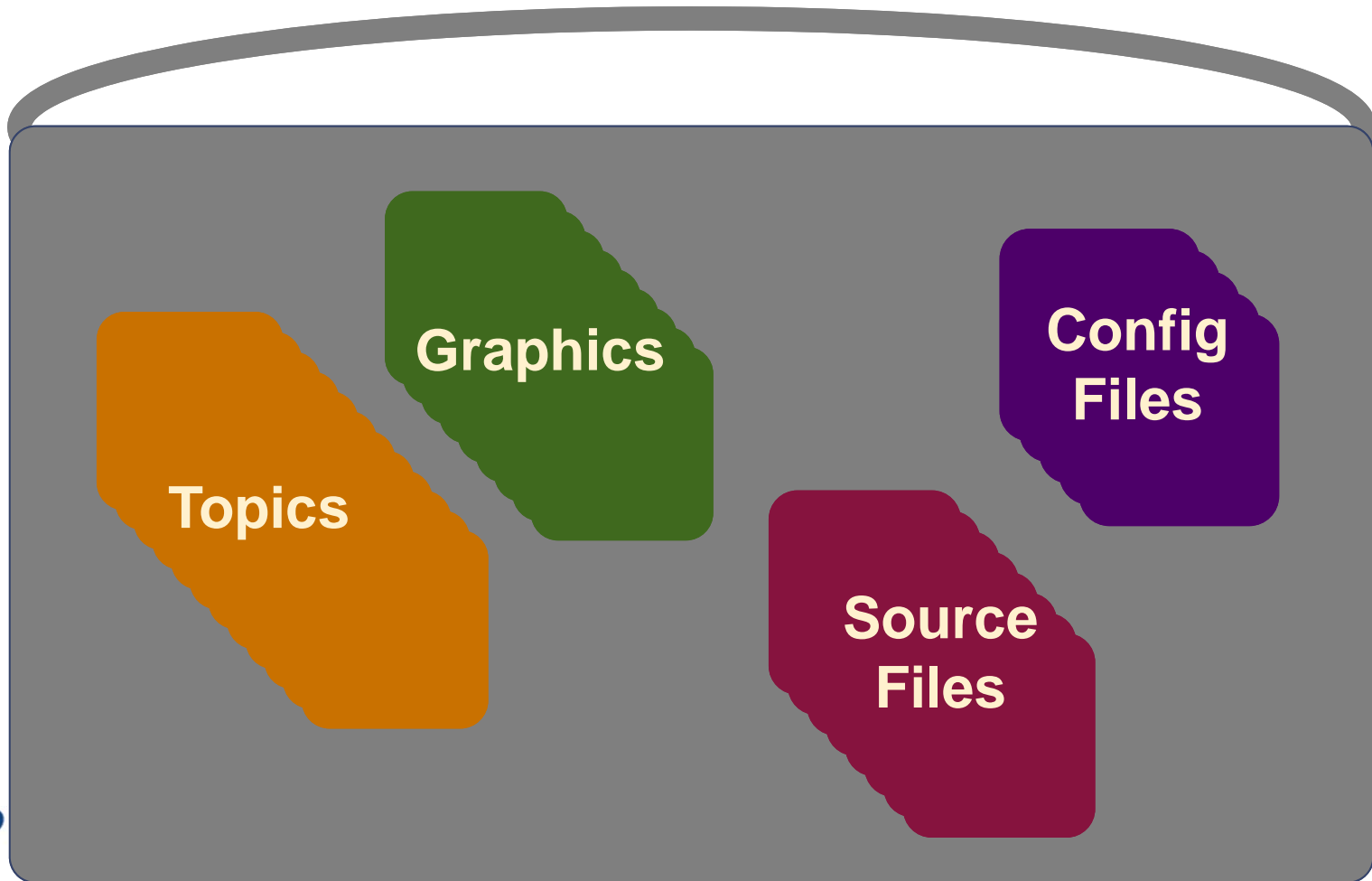
How to Develop an Easily Understood Metric ... Step 1

- Define the components we are measuring
 - What is a **REPOSITORY**?
 - What is a **SUITE**?
 - What is a **RELEASE**?
 - What do writers **DO**?

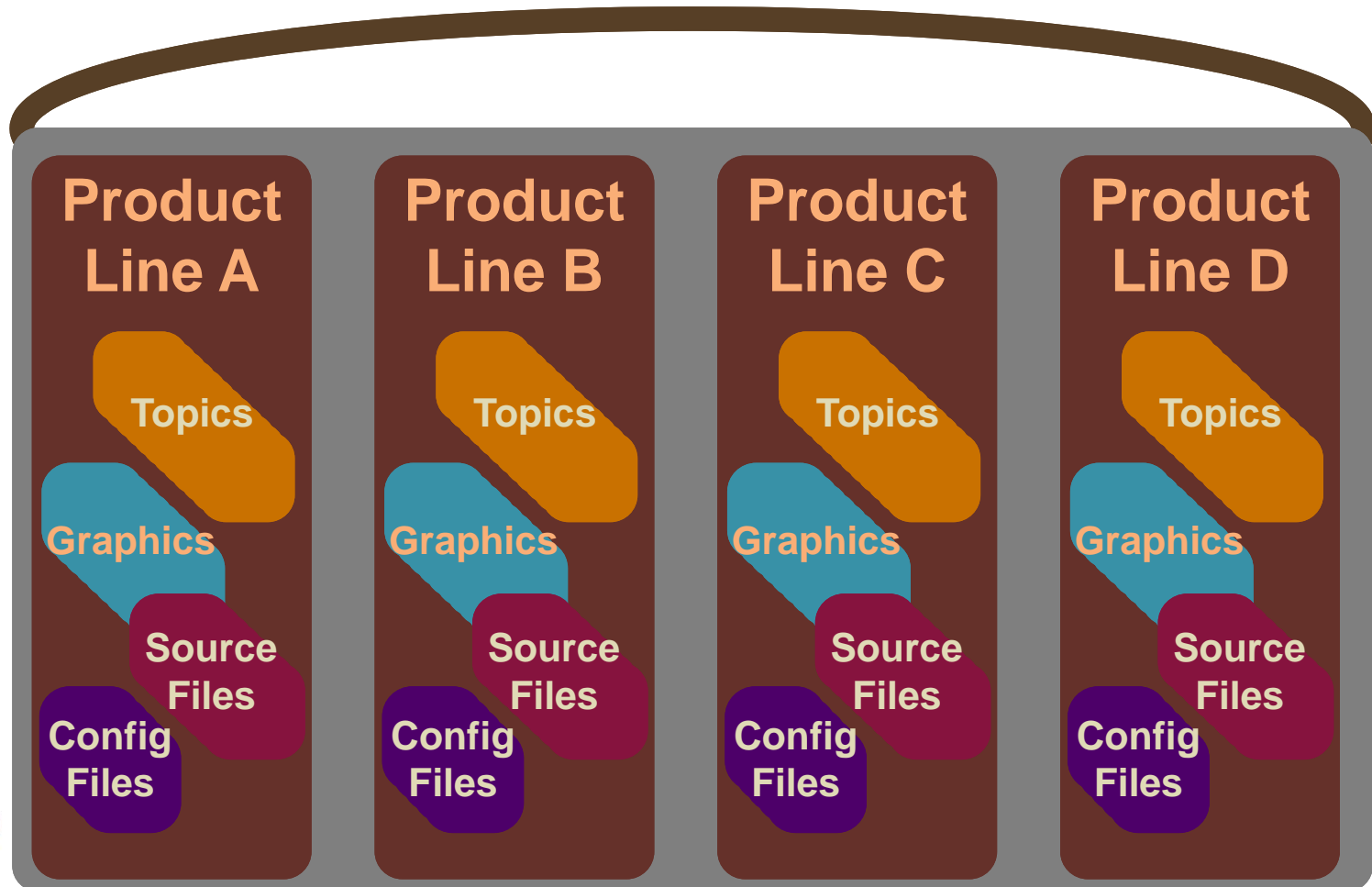


What is a REPOSITORY?

It's a Bucket



What is a REPOSITORY? Or Maybe ... a Bucket of Buckets?



What is a REPOSITORY?

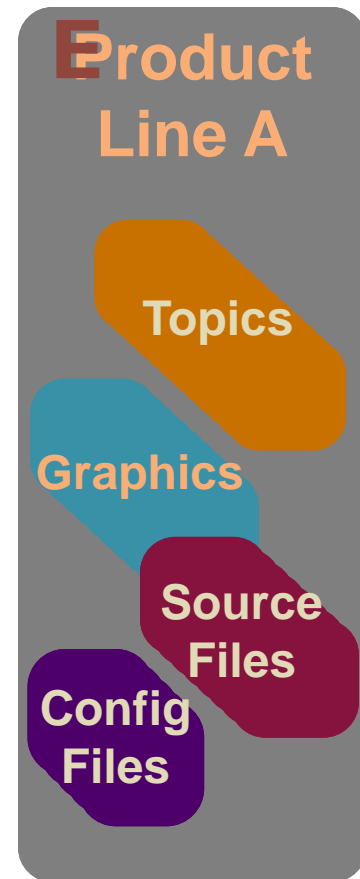
- REPOSITORY is not well defined across groups and companies.
- Not well defined in terms of business processes outside of documentation.
- **We can IGNORE REPOSITORY**



What is a SUITE?

- A collection of information that is used for publishing a deliverable
- A SUITE is related to:
 - Budget
 - Project
 - Workload
 - Product, Products, Product Line

SUIT



What is a RELEASE?

- **ONE** publishing event of an entire **SUITE** delivered to customers.
- If content appears in more than one output (for example PDF + Help), **always count that content ONLY ONCE.**



What do writers DO?

- Without DITA:
 - They write, and
 - Fix layout mostly and write, and
 - Try to publish and write, and
 - Fiddle with tools etc
- **With DITA:**
 - **Mostly they write ... words**



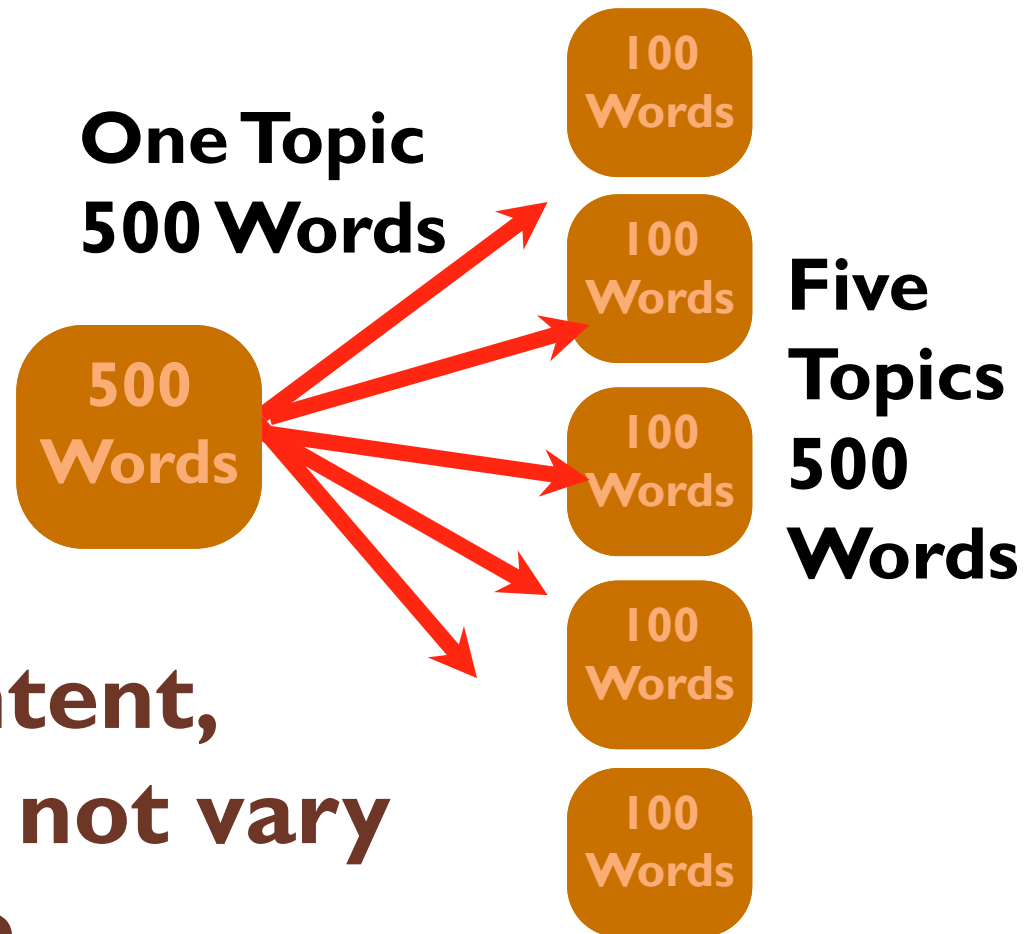
Our Easy-to-Understand Metric is Based On ...

- SUITE?
 - **Information Supporting a Product(s)**
- RELEASE?
 - **A SUITE Publishing Event = Deliverable**
- What do writers DO?
 - **They Write Words**



Why not use Topics or a Similar Concept ?

- **TOPICS vary over time.**
- **For the same content, words do not vary over time.**



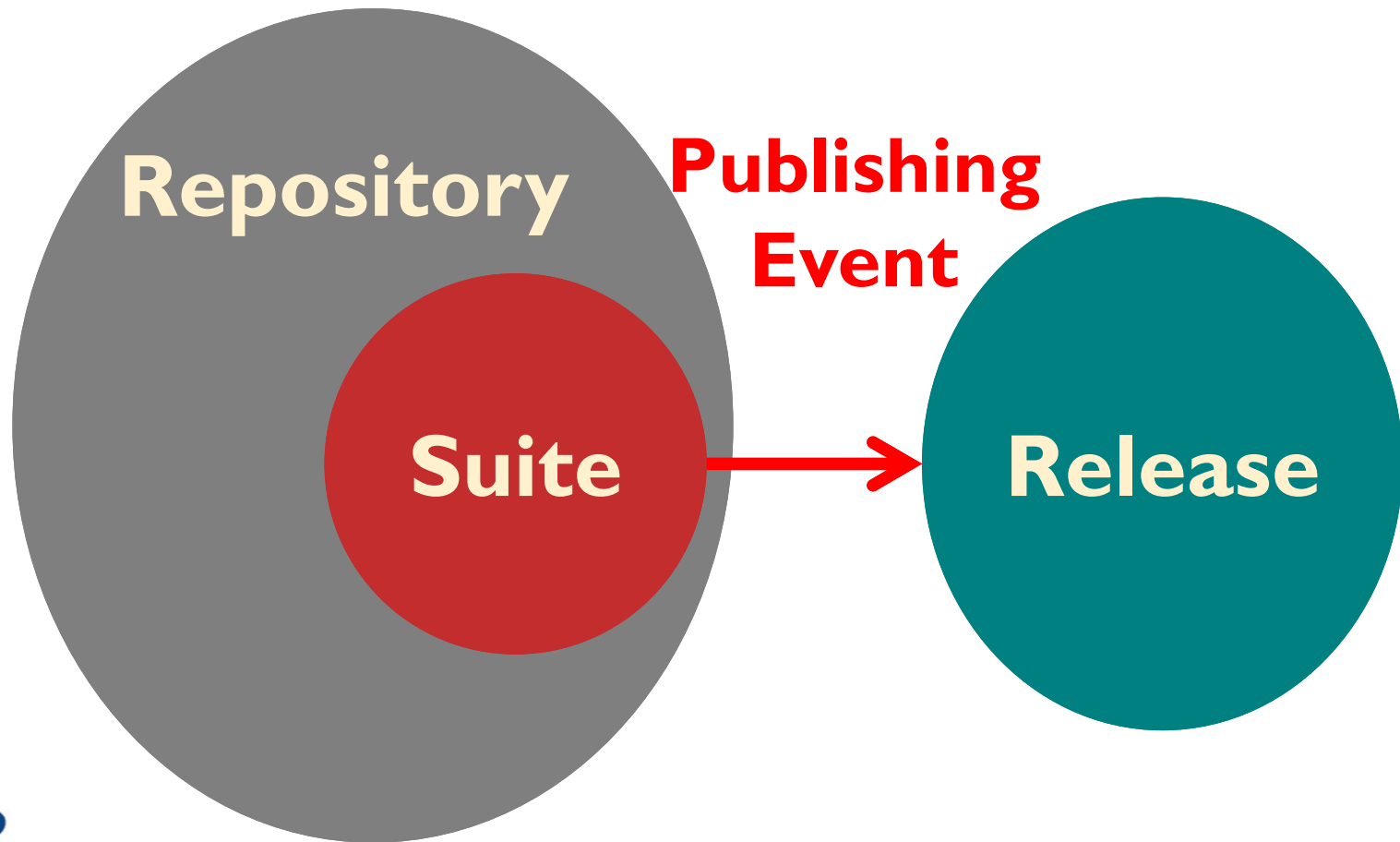
SAMALANDER



BASELINE FOR TECHNOLOGY AND TIME INDEPENDENT METRICS

Apple-to-Apple Metrics for DITA: CIDM Content Management Strategies / DITA Conference,
Rhode Island, 2013-04-15-17

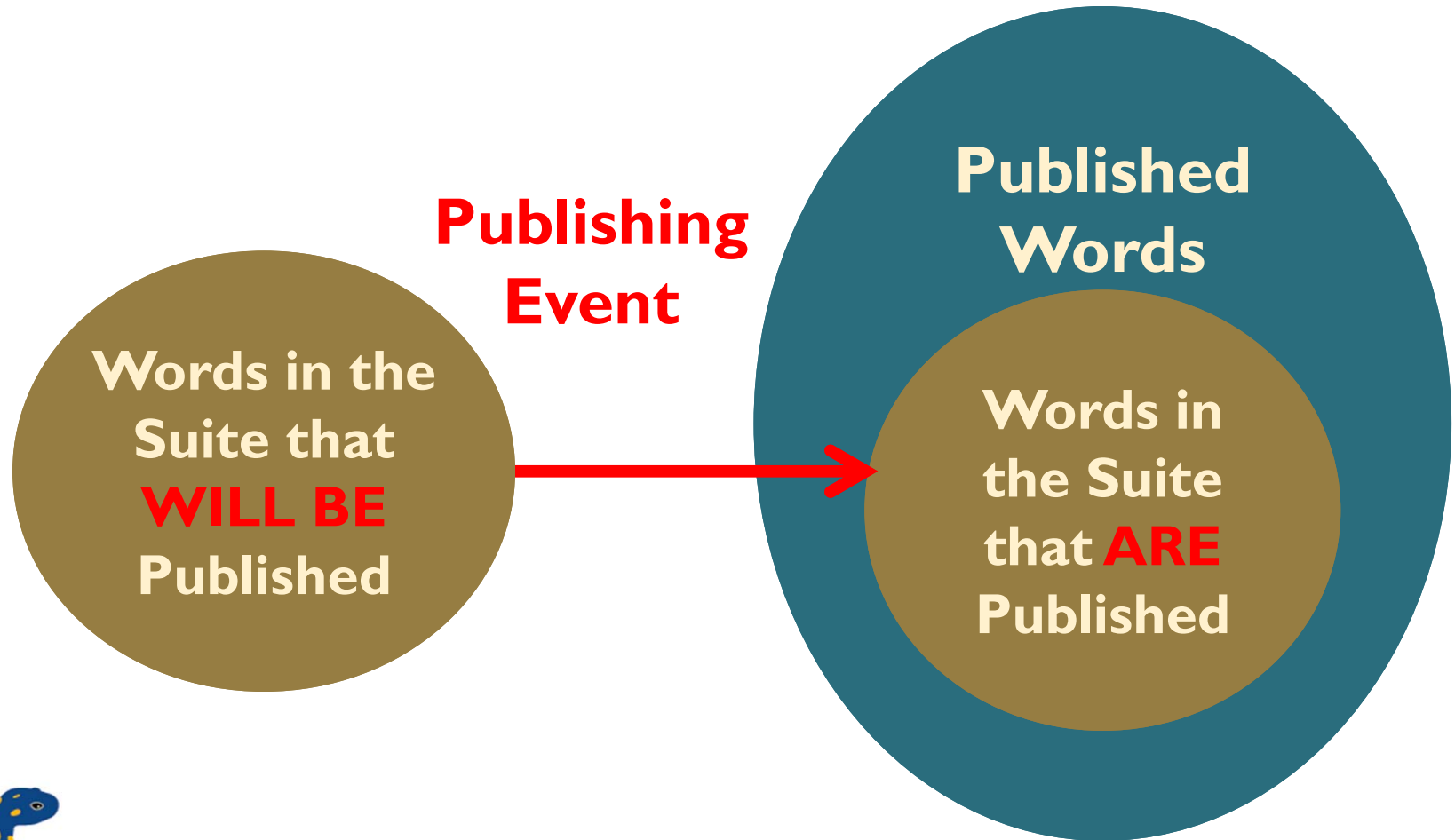
Now for the Math ...



Count What is Published ... Ignore Repository



We can also Ignore the SUITE



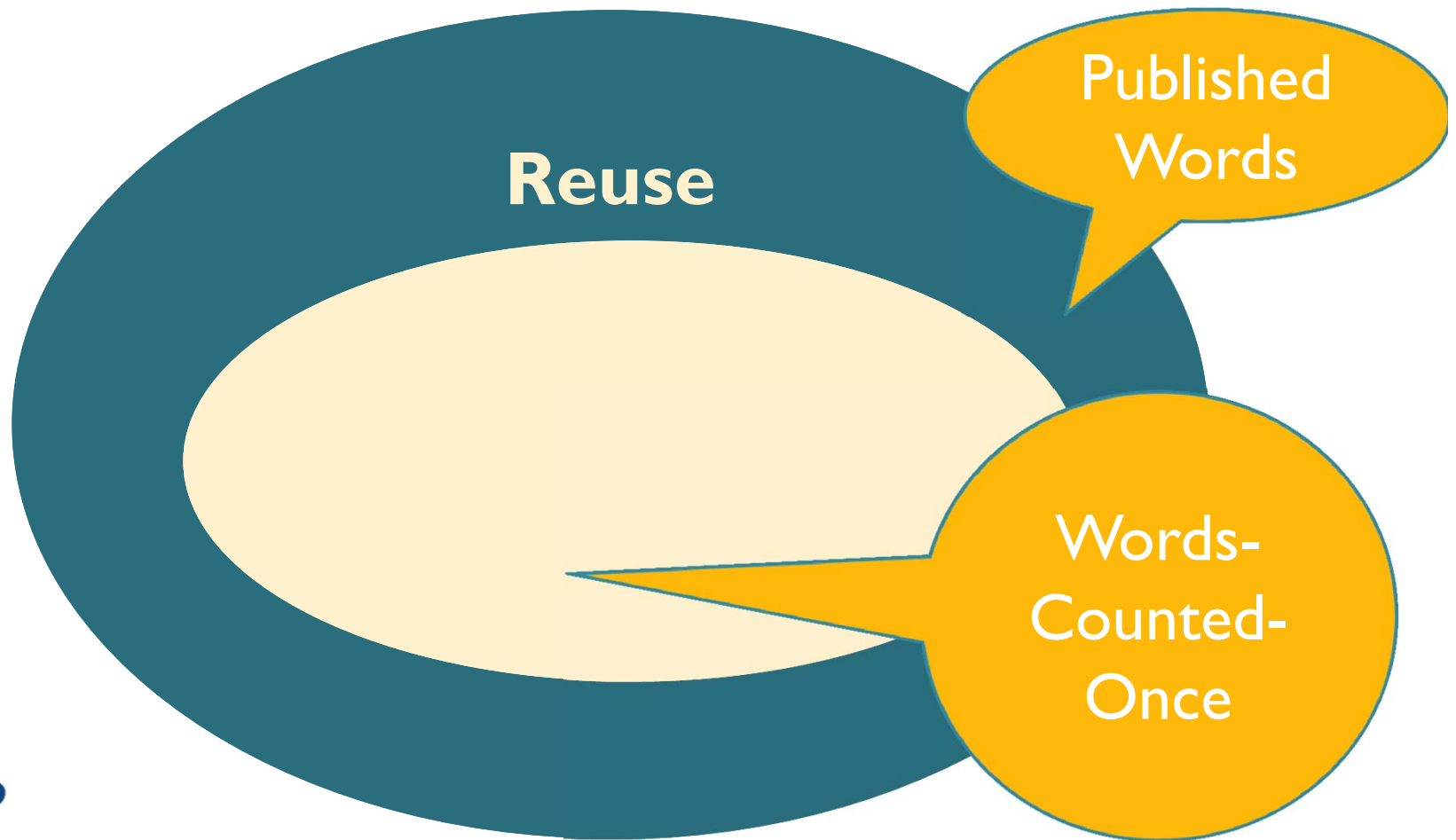
We are Left With **Published Words** and **Words-Counted-Once**

Published Words

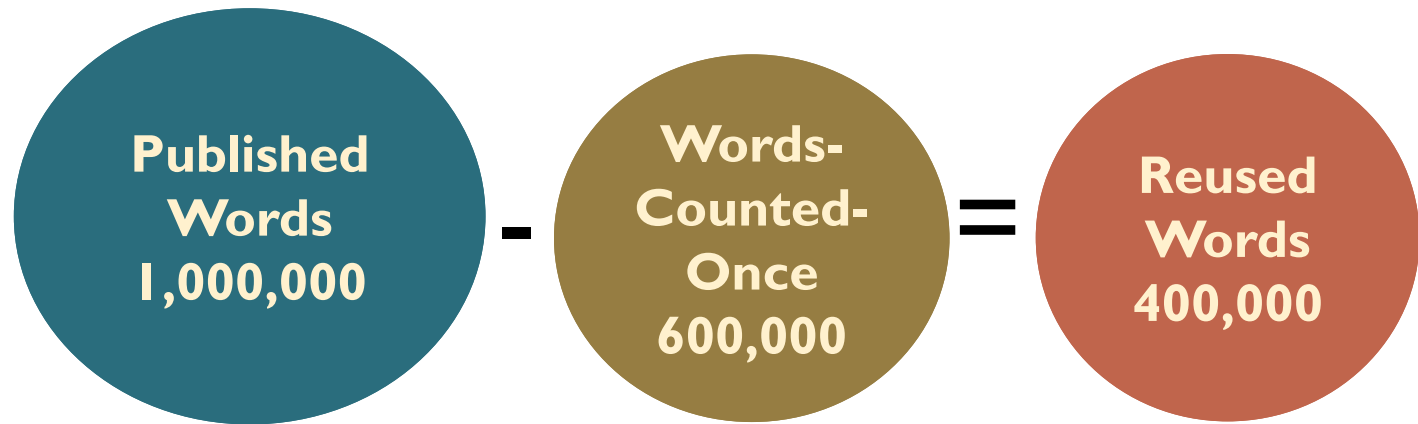
Words in the Suite that
ARE Published =
**Words-Counted-
Once**



Reused Words



Reuse is Defined as ...



$$\left(1 - \left(\frac{\text{Words-Counted-Once}}{\text{Published Words}} \right) \right) \times 100 = \text{Reused \%}$$

$$\left(1 - \left(\frac{600,000}{1,000,000} \right) \right) \times 100 = 40 \%$$



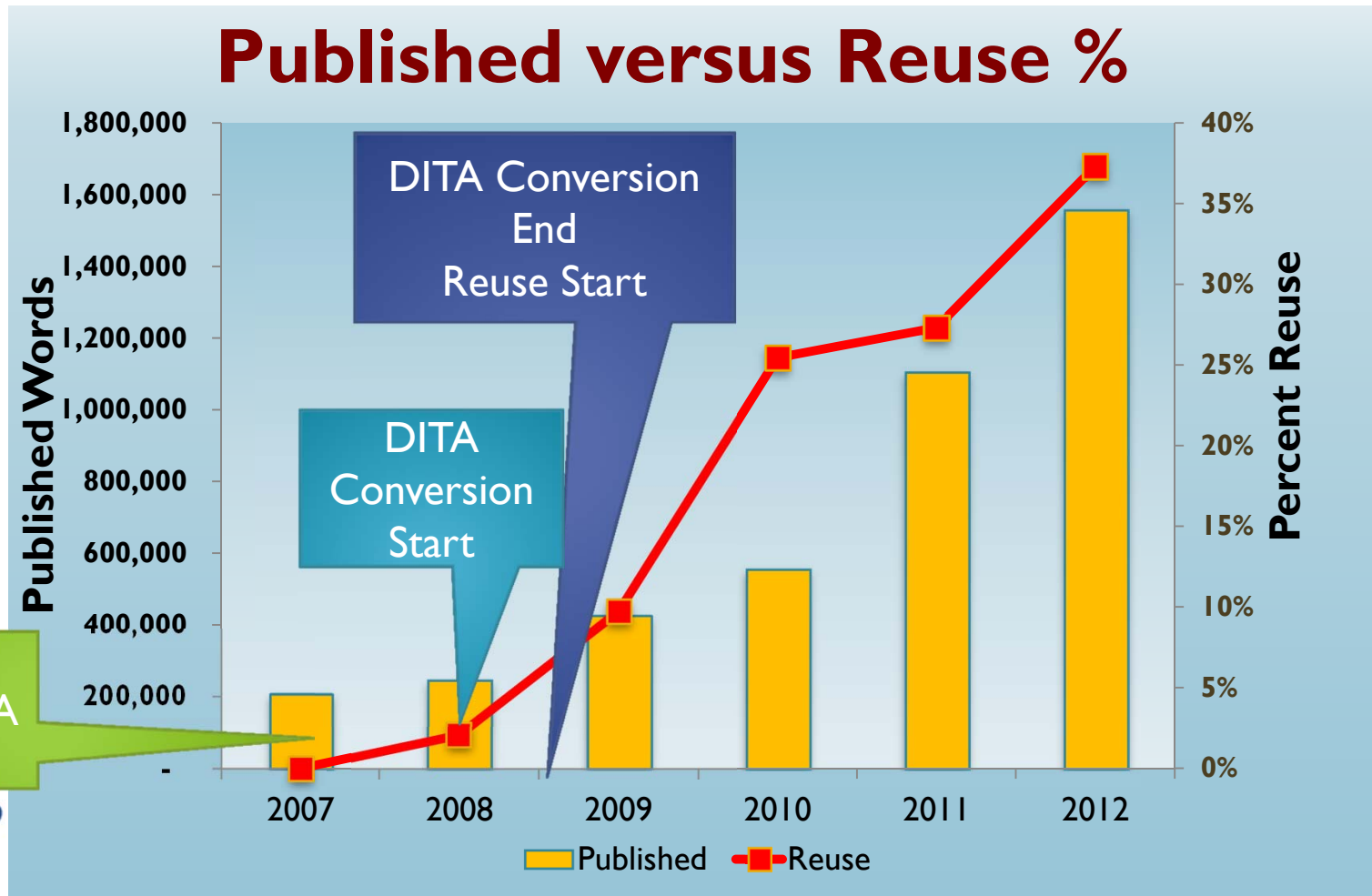
SAMALANDER



EXAMPLES

Actual Example

Reuse and Published



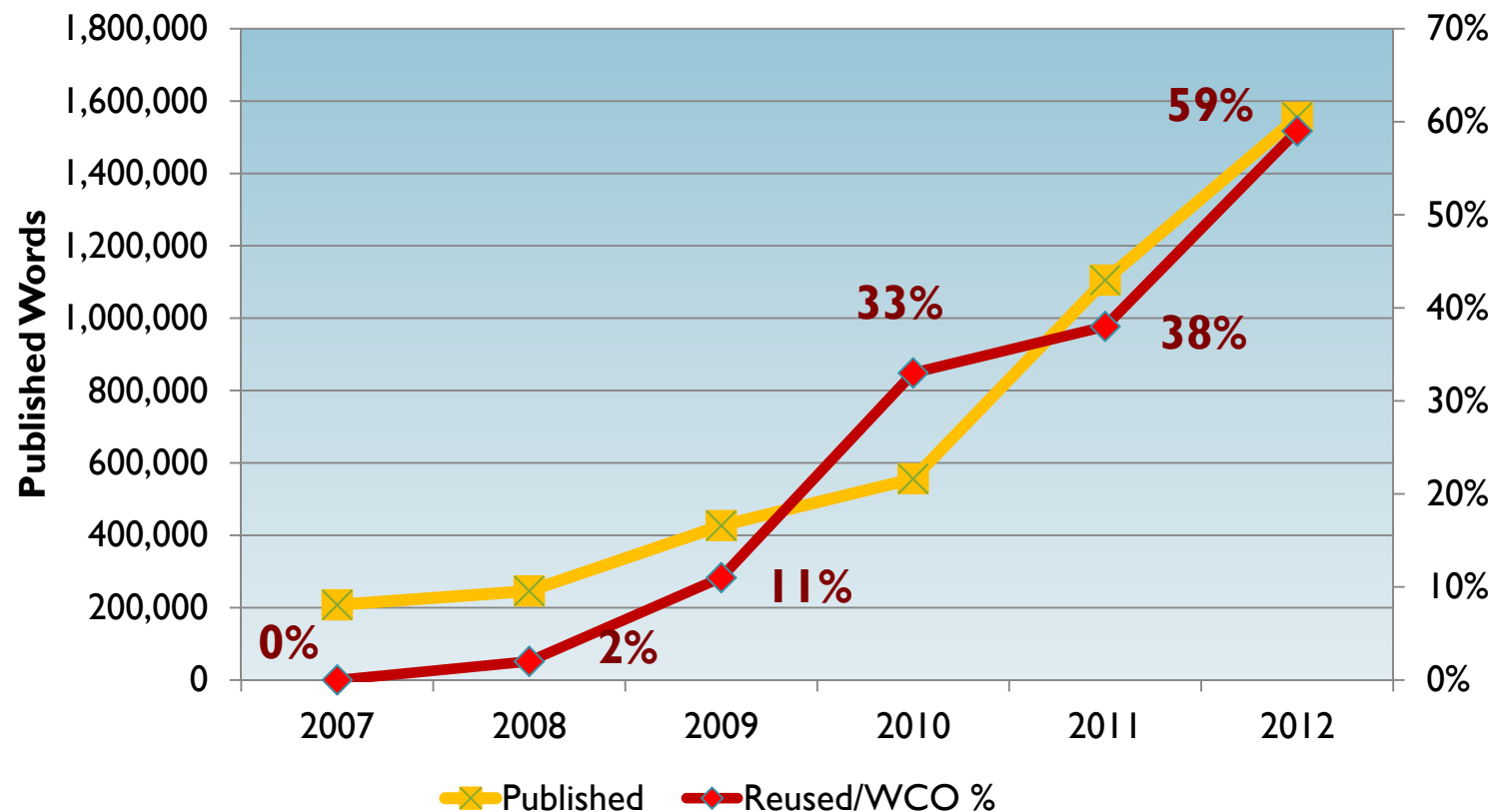
Before DITA



Actual Example: How it Works

...

Published Words vs Reused / Words-Counted-Once



Apple-to-Apple Metrics for DITA: CIDM Content Management Strategies / DITA Conference, Rhode Island, 2013-04-15-17



Example: ROI Based on Published vs Words-Counted-Once

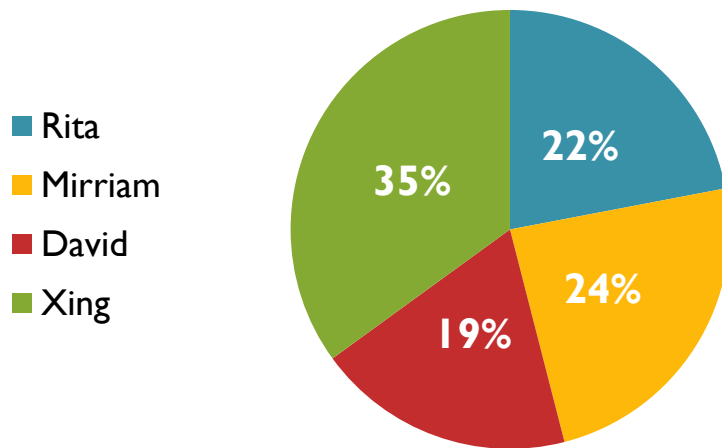
| Year | Incremental Cost | Published Words in Suite(s) | Published Words % Change over 2007 | Savings in-year Published | Words-Counted-Once in Suite(s) | Words-Counted-Once % Change over 2007 | Savings in-year Words-Counted-Once |
|------|------------------|-----------------------------|------------------------------------|---------------------------|--------------------------------|---------------------------------------|------------------------------------|
| 2007 | \$15,000 | 207,450 | - | \$ - | 207,450 | - | \$ - |
| 2008 | 0 | 246,150 | 18.7% | \$ 3,655 | 241,000 | 16.2% | \$ 1,173 |
| 2009 | 0 | 426,913 | 105.8% | \$ 105,791 | 392,239 | 89.1% | \$ 89,076 |
| 2010 | 0 | 556,116 | 168.1% | \$ 168,072 | 434,520 | 109.5% | \$ 109,458 |
| 2011 | \$400 | 1,105,067 | 432.7% | \$ 432,291 | 803,277 | 287.2% | \$ 286,815 |
| 2012 | 0 | 1,554,802 | 649.5% | \$ 649,483 | 974,872 | 369.9% | \$ 369,931 |

Note: Savings calculated based on Loaded Labor Rate of \$100,000/year

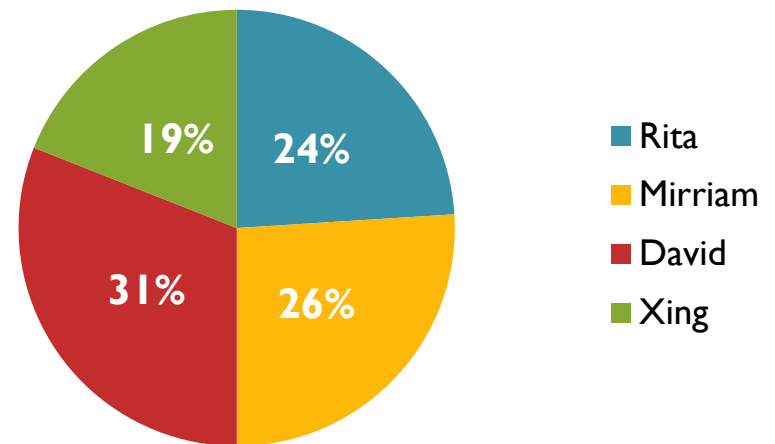


Example: Metrics by Author

Author Workload in Published Words



Author Workload in Words-Counted-Once



Example: Weighting Topics by Usage in a Release

| Topic | Words in Topic |
|----------|----------------|
| ABC.dita | 50 |
| GHI.dita | 250 |
| STU.dita | 5000 |
| JKL.dita | 500 |
| PQR.dita | 2000 |
| DEF.dita | 100 |
| MNO.dita | 1000 |



Example: Weighting Topics by Usage in a Release

| Topic | Words in Topic | Number of Times Used | Reused Value in Suite | Editorial Priority |
|--------------|----------------|----------------------|-----------------------|--------------------|
| ABC.dita | 50 | 150.00 | 7,500 | 1 |
| GHI.dita | 250 | 25.00 | 6,250 | 2 |
| STU.dita | 5,000 | 1.00 | 5,000 | 3 |
| JKL.dita | 500 | 5.00 | 2,500 | 4 |
| PQR.dita | 2,000 | 1.00 | 2,000 | 5 |
| DEF.dita | 100 | 10.00 | 1,000 | 6 |
| MNO.dita | 1,000 | 1.00 | 1,000 | 6 |
| Total | 8,900 | | 25,250 | 6 |



Example: Weighting Topics by Usage and Translation

| Topic | Words in Topic | Number of Times Used | Reused Value in Suite | Translation Priority = Editorial Priority | Translation Cost with Reuse @ \$0.25/word | Translation Cost Without Reuse @ \$0.25/word | Savings | Savings % |
|--------------|----------------|----------------------|-----------------------|---|---|--|----------------|------------|
| ABC.dita | 50 | 150 | 7,500 | 1 | \$13 | \$1,875 | \$1,862 | 99% |
| GHI.dita | 250 | 25 | 6,250 | 2 | \$63 | \$1,563 | \$1,500 | 96% |
| STU.dita | 5,000 | 1 | 5,000 | 3 | \$1,250 | \$1,250 | \$0 | 0% |
| JKL.dita | 500 | 5 | 2,500 | 4 | \$125 | \$625 | \$500 | 80% |
| PQR.dita | 2,000 | 1 | 2,000 | 5 | \$500 | \$500 | \$0 | 0% |
| DEF.dita | 100 | 10 | 1,000 | 6 | \$25 | \$250 | \$225 | 90% |
| MNO.dita | 1,000 | 1 | 1,000 | 6 | \$250 | \$250 | \$0 | 0% |
| Total | 8,900 | | 25,250 | | \$2,225 | \$6,313 | \$4,088 | 65% |



How does Words-Counted-Once Compare to “DITA Metrics 101”

- “DITA Metrics 101” by Mark Lewis.
- Latest book on DITA Metrics
- What are the differences???



“DITA Metrics 101”

“DITA 101”: Cost of a Simple Project with Reuse

| Product 1 | Number of Topics | Cost per Topic | Percent Reuse | Without DITA (hrs) | With DITA (hrs) | Savings (hrs) |
|--------------------------|------------------|----------------|---------------|--------------------|-----------------|---------------|
| Task | 65 | 5 | 0.5 | 325 | 162.5 | 162.5 |
| Concept | 25 | 3 | 0.9 | 75 | 7.5 | 67.5 |
| Reference | 10 | 4 | 0.25 | 40 | 30 | 10 |
| Product 1 Total | 100 | | | 440 | 200 | 240 |
| Suite Total (3 Products) | | | | 1,320 | 600 | 720 |

Estimated.
Stable over
Time?



“Words-Counted-Once”

Words-Counted-Once: Cost of a Simple Project with Reuse

| Product 1 | Number of Topics | Hours / 500 Words | Published Words | Words Counted Once | Percent Reuse | Without DITA (hrs) | With DITA (hrs) |
|--------------------------|------------------|-------------------|-----------------|--------------------|---------------|--------------------|-----------------|
| Task | 65 | 5 | 32,500 | 16,250 | 0.5 | 325 | 162.5 |
| Concept | 25 | 3 | 12,500 | 1,250 | 0.9 | 75 | 7.5 |
| Reference | 10 | 4 | 5,000 | 3,750 | 0.25 | 40 | 30 |
| Product 1 Total | 100 | | | | | 440 | |
| Products 2 and 3 | 200 | | | | | 880 | 400 |
| Suite Total (3 Products) | | | | | | 1320 | 840 |

Measured.
Stable over
Time.



Difficulty Weighting and Words-Counted-Once: Increased Accuracy

- ```
<task id="My-Best-Task-22AEF">
 <title>Topic title</title>
 <prolog>
 <author>
 Martha
 </author>
 <metadata>
 <othermeta name="effort-type" content="new"/>
 <!-- "new" OR "revised" OR "touched" OR "untouched"-->
 <othermeta name="effort-prod-class" content="hardware"/>
 <!-- "hardware" OR "software">
 <othermeta name="effort-complexity" content="complex"/>
 <!-- "complex" OR "normal" OR "simple" -->
 </metadata>
 </prolog>
 <taskbody>
 <p>Words-Counted-Once</p>
 </body>
```





# REFERENCES: Metrics Module

---

- Download Metrics Module:

<http://samalander.com>

- So far it works with DITA on disk ... see release notes.
- Take our survey please: what would you like to see next?

