



I18N A1D D5T: W1O D2S W2T?

Internationalization and DITA-OT: Who Does What?

Robert D. Anderson, IBM

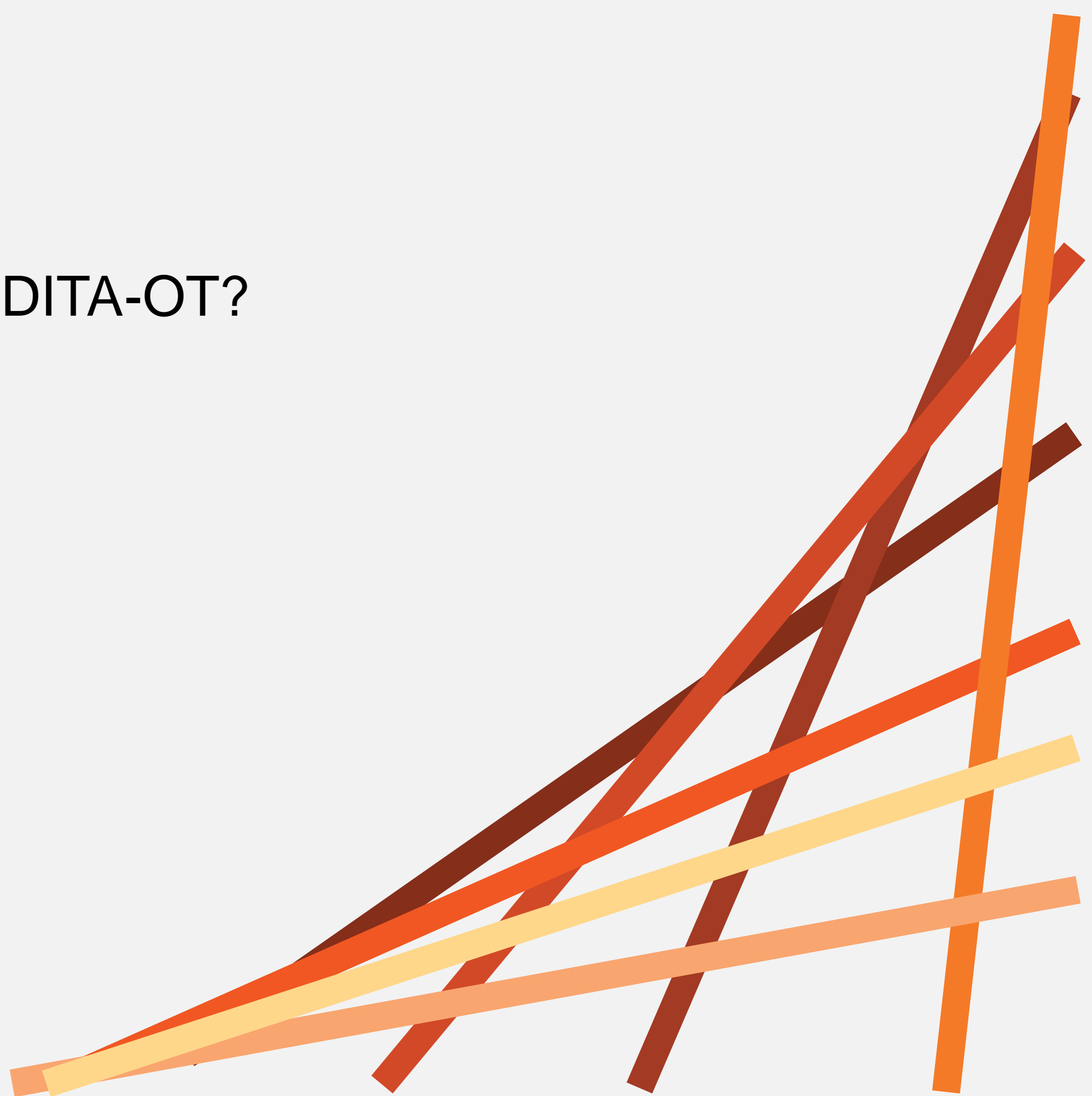
@robander

November 13, 2016



Agenda

- Why are we here?
- Internationalization: what is the role of DITA-OT?
- Where does your part come in?
- Should you be scared?
- Scary story break
- No, really – should you be scared?



What I expect you are expecting...

- I18N: What does DITA-OT do for you?
- I18N: What's left for you?
- I18N: Where do localization vendors or translators come in?



- This is an introductory talk. *But it is not:*
 - *A deep dive into machine translation*
 - *The one and only answer to translation and DITA*



What am I missing?

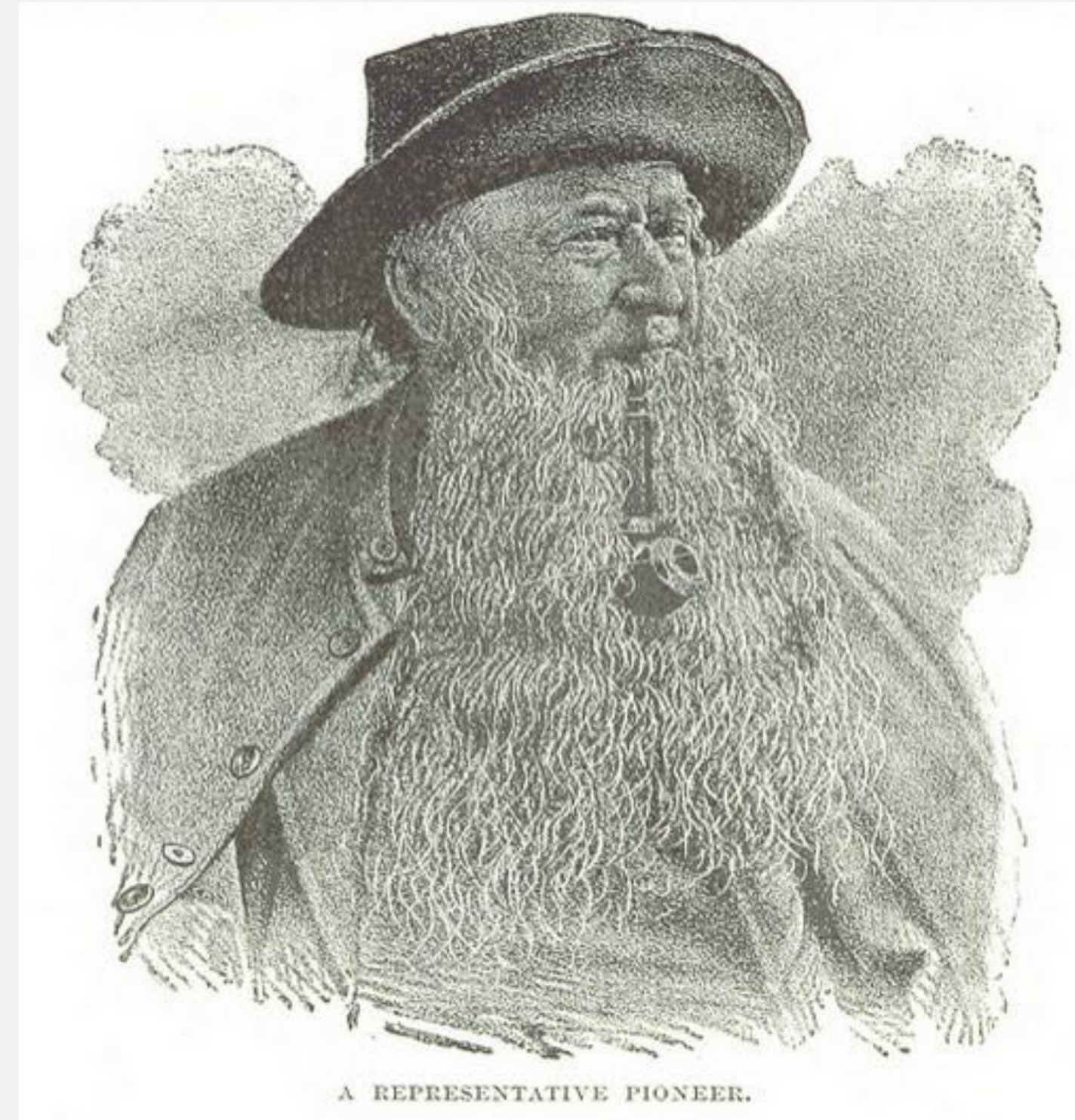
What should I expect...?

- In the audience...
 - New to DITA-OT?
 - Already went through translation cycle?
 - In a CMS, or file system?
 - Already know everything I'm going to say?
- I expect you to ask questions when you have them
- *Share your own experiences!*



Why am I the one up here?

1. I've been working with the DITA and the toolkit since *The Beginning* (of those things)
2. I like languages and talking about them
(*I speak approximately 2 and a quarter*)
3. We (IBM) use DITA-OT processing to publish DITA into multiple formats in ~50 languages
4. There's still a lot of stuff I don't know



What is DITA-OT's role?

THE IDEAL:

*You should be able to publish localized content
just as easily in one language as another*



Supported languages and scripts

- 48 for PDF and HTML (including 2 scripts for Serbian)
- Supporting a language in DITA-OT (definition): If you have a topic or map in language X, out-of-the-box support should be equivalent to the default English*

* *As much as possible given support and license constraints*

<http://www.dita-ot.org/2.4/user-guide/DITA-globalization-xhtml.html>

<http://www.dita-ot.org/2.4/user-guide/DITA-globalization-pdf.html>

А а (A)	Р р (R)
Б б (B)	С с (S)
В в (V)	Т т (T)
Г г (G)	У у (U)
Д д (D)	Ф ф (F)
Е е (E)	Х х (KH)
Ё ё (YO)	Ц ц (TS)
Ж ж (ZH)	Ч ч (CH)
З з (Z)	Ш ш (SH)
И и (I)	Щ щ (SHCH)
Й й (Y)	Ъ ъ (-)
К к (K)	Ы ы (Y)
Л л (L)	Ь ь (')
М м (M)	Э э (E)
Н н (N)	Ю ю (YU or IU)
О о (O)	Я я (YA or IA)
П п (P)	

How does I18N-ized content differ?

$a \neq b$

- Fonts
- Generated text
- Other:
 - Spacing / hyphenation
 - Index sorting
 - Directionality
 - (Possibly) Ease with which you, personally, can read it
 - *What else? Bonus points for anything I've missed*

Fonts

... are complex.

Don't believe me? Check out [Just My Type: A Book About Fonts](#) by Simon Garfield...

- HTML: Use CSS to change.
- PDF: Default sets fonts by character. Add new mappings to change. Or ...
- ... “Font configuration in PDF2”
<http://www.elovirta.com/2016/02/18/font-configuration-in-pdf2.html>

Generated text or headings

From <http://www.dita-ot.org/dev/user-guide/build-using-dita-command.html>



What to do next

Usually, you will want to specify a set of reusable build parameters in a `.properties` file.

- **Setting build parameters with `.properties` files**

Usually, DITA builds require setting a number of parameters that do not change frequently. You can reference a set of build parameters defined in a `.properties` file when building output with the `dita` command. If needed, you can override any parameter by specifying it explicitly as an argument to the `dita` command.

- **Migrating Ant builds to use the `dita` command**

Although Ant builds are still supported in the DITA Open Toolkit, you might want to switch to use the new `dita` command.

Related reference

[Arguments and options for the `dita` command](#)

[DITA-OT parameters](#)

[Internal Ant properties](#)

Other items handled by DITA-OT

- Index sorting:
 - DITA-OT 2.3 handles PDF index collation, headings
 - Exceptions: Japanese, Chinese (licensed alternatives available)
- Directionality:
 - DITA-OT 2.3 defaults to Right-to-left for predetermined languages
 - `@dir` attribute in the text is respected



Before you push the “translate” button

*Reminder: DITA-OT does not translate your content
(Unless Jarno has plans for an incredible premium edition)*



Lots of best practices for writing for translation

... which I'm not going to cover here. I'm going for technical DITA-OT stuff.

Except for this, because ugh, don't do this:

`<p>Updates to your <keyword conref="#./thing"/>s are required in the following circumstances: ...</p>`

- English: Works as long as “thing” resolves to “routine”.
- English: **Fails when “routine” changes to “process”**
- **`<ye11>Fails when you translate to most languages</ye11>`**

Assuming the content is correct, you're left with...

- Font preferences
 - Use CSS, or use PDF font configuration
- Generated text: need “Note” to be “Hey, look at this!!!!”?
 - Extend with your own strings
 - Most localization vendors expect this
 - Tip: put new natural language strings in variable files, not in stylesheets
- Hyphenation: some languages might need this; not all formatters support
 - We extend our PDF output to turn it on for select languages

** All of these apply to the original language!*

What if you need Welsh (or something else)?

1. Translate the default strings in common and (if needed) PDF directory; add language value to string mapping file

Tip: Use a plug-in to ensure you can upgrade to new releases

(Two directories, there for historical reasons, may be consolidated in the future)

2. If index sorting is needed, create collation file using existing PDF code as a model

This can be simple, or very difficult, depending on the language.

3. You may need new fonts if using a new script

4. *Please: contribute the new support back to DITA-OT!*



Caveat: IBM's localization process



Localization vendor's role (within DITA-OT context)

- Translate content (including strings as needed)
- Validating output
- Some tools require 100% markup equivalence between source / target
- Most CMSs store your translated content

- *Bonus point time: what am I missing, particularly in the context of DITA-OT and the publishing process?*



Should you be scared of I18N?

No.

Nein.

Нет.

Nu.

Non.

Nej.

Ei.

Nei.

Không.

Үгүй.

...

But what about those stories...



From: https://en.wikipedia.org/wiki/Tucson,_Arizona#/media/File:Saguaro_Sunset.jpg

License: <https://creativecommons.org/licenses/by/3.0/>

Caveat: the call could easily have come from anywhere else

Quote from anonymous IBMer

“In most people's heads, DEALING WITH TRANSLATION equates to the Sword of Damocles.

It's scary, it's hard, I'm going to get errors I don't understand, it's not going to work out, I'll have to redo it three times because someone yelled at me about some other error that makes no sense to me, I don't understand what 90% of this process is, and oh, by the way, keep working on the new features that are coming out so you can do this again in 90 days.”

November 8, 2016 (last week, completely unprompted)

No, really. Should you be scared?

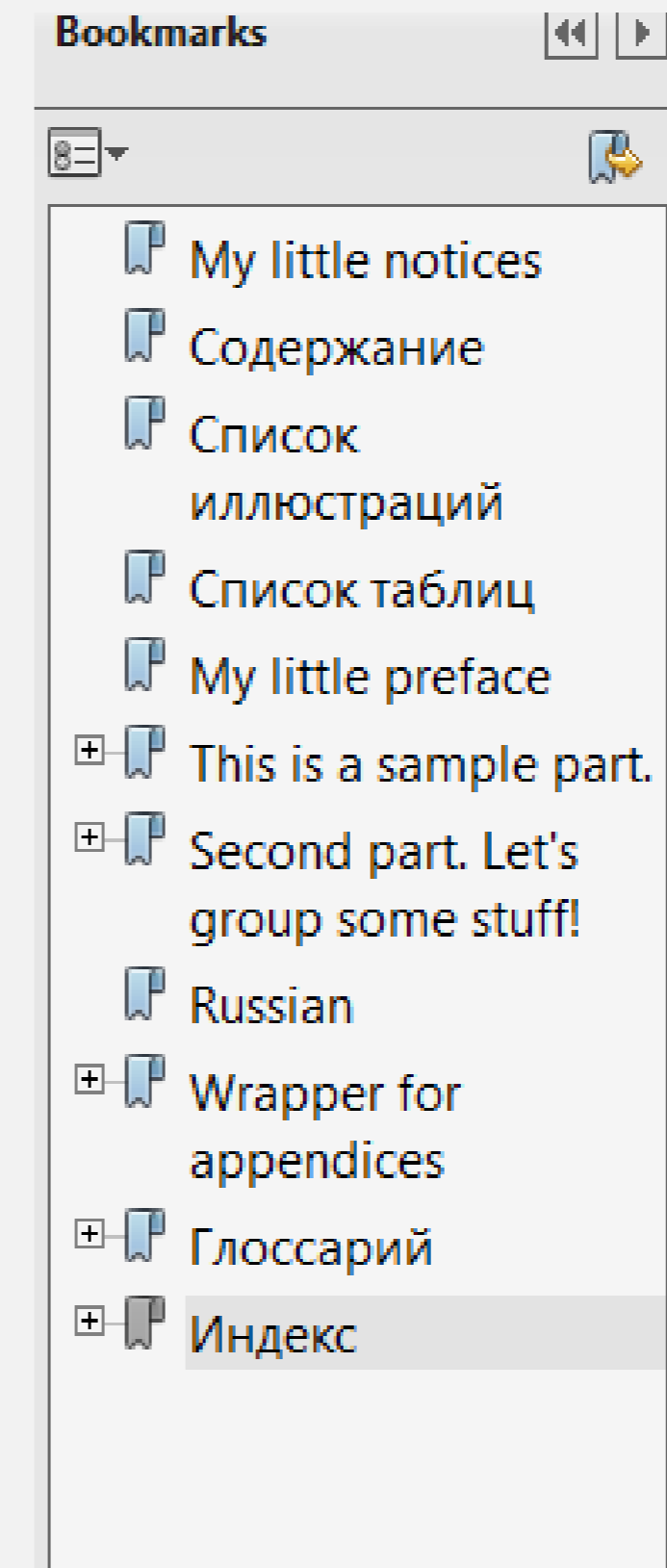
No.

But I'm happy to hear more horror stories if you have them!



Possibly interesting samples

- Set of NLS test files for all supported languages:
<https://github.com/robander/metadita/tree/master/sampledocs/NLStext>
- Tests generated strings + index collation
- Covers most languages & characters;
Chinese, Japanese, Korean are incomplete
- Also includes (unsupported) Welsh test



Индекс

Спец. символы

! exclamation point (1) 31
? question (12) 32
. period (13) 32
(left paren (9) 32
) right paren (16) 32
[left bracket (11) 32
{ left curly brace (10) 32

DITA-OT: Further study

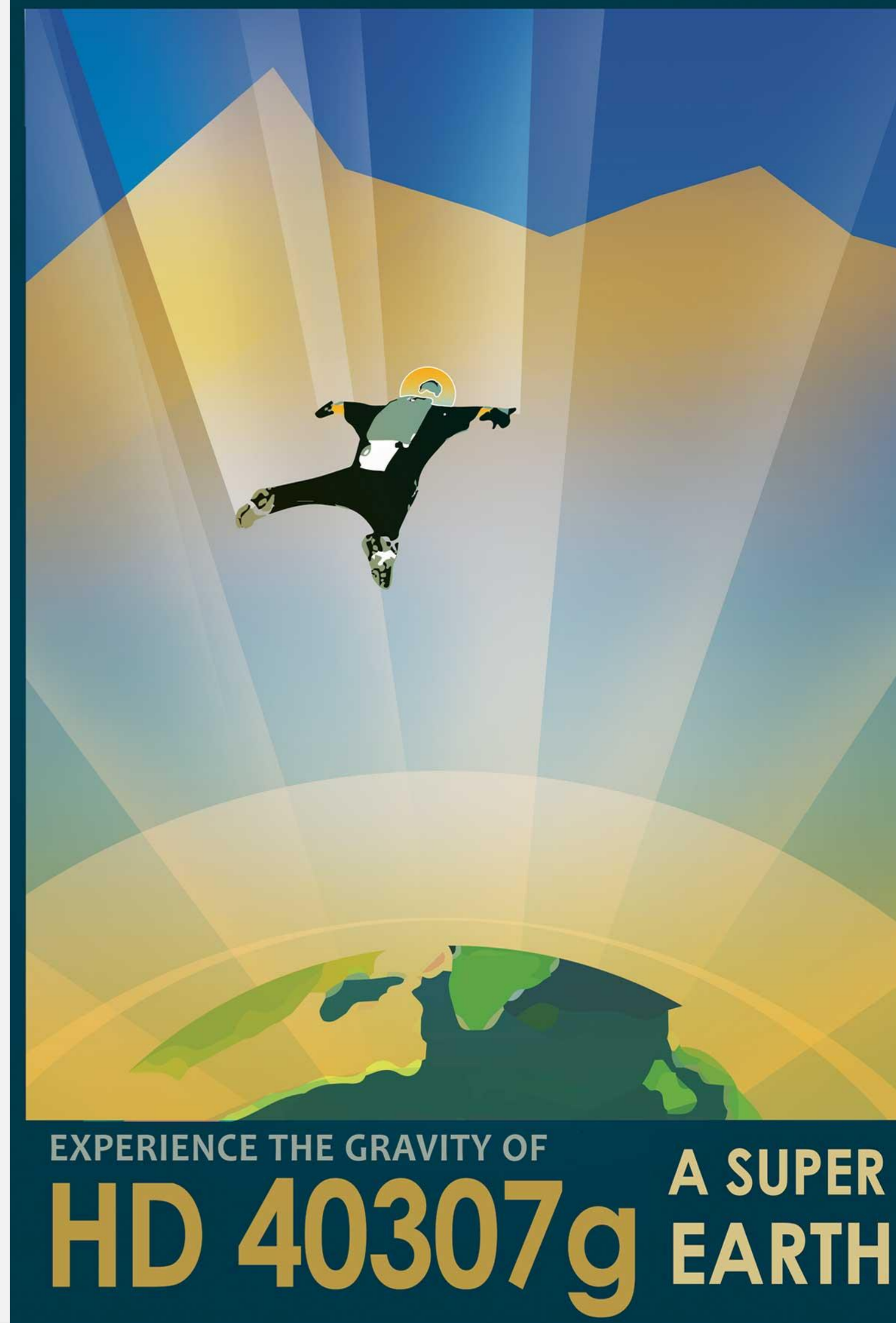
Monthly DITA-OT Contributor calls: hosted by Syncro Soft, open to anyone

Monthly DITA-OT Docs calls: hosted by Eberlein Consulting, open to anyone

Github project: <https://github.com/dita-ot/dita-ot/>

Everything else at <http://dita-ot.org>

Get involved! Please!



Super secret DITA-OT shortcut resources!

Maybe these are useful? If so, tell your friends!

If you'd like to keep your friends, just tell your co-workers!

- <http://code.dita-ot.org> → redirects to <http://github.com/dita-ot/dita-ot/>
- <http://issues.dita-ot.org> → redirects to <http://github.com/dita-ot/dita-ot/issues/>
- <http://wiki.dita-ot.org> → redirects to <http://github.com/dita-ot/dita-ot/wiki>
(Used for contributor meeting minutes)

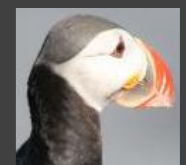
Image credits

Jetpack Image Courtesy NASA/JPL-Caltech

<http://www.jpl.nasa.gov/visions-of-the-future/>

Old-time images from British Library Flickr stream

www.flickr.com/photos/britishlibrary/



Thank you

Digital Services Group