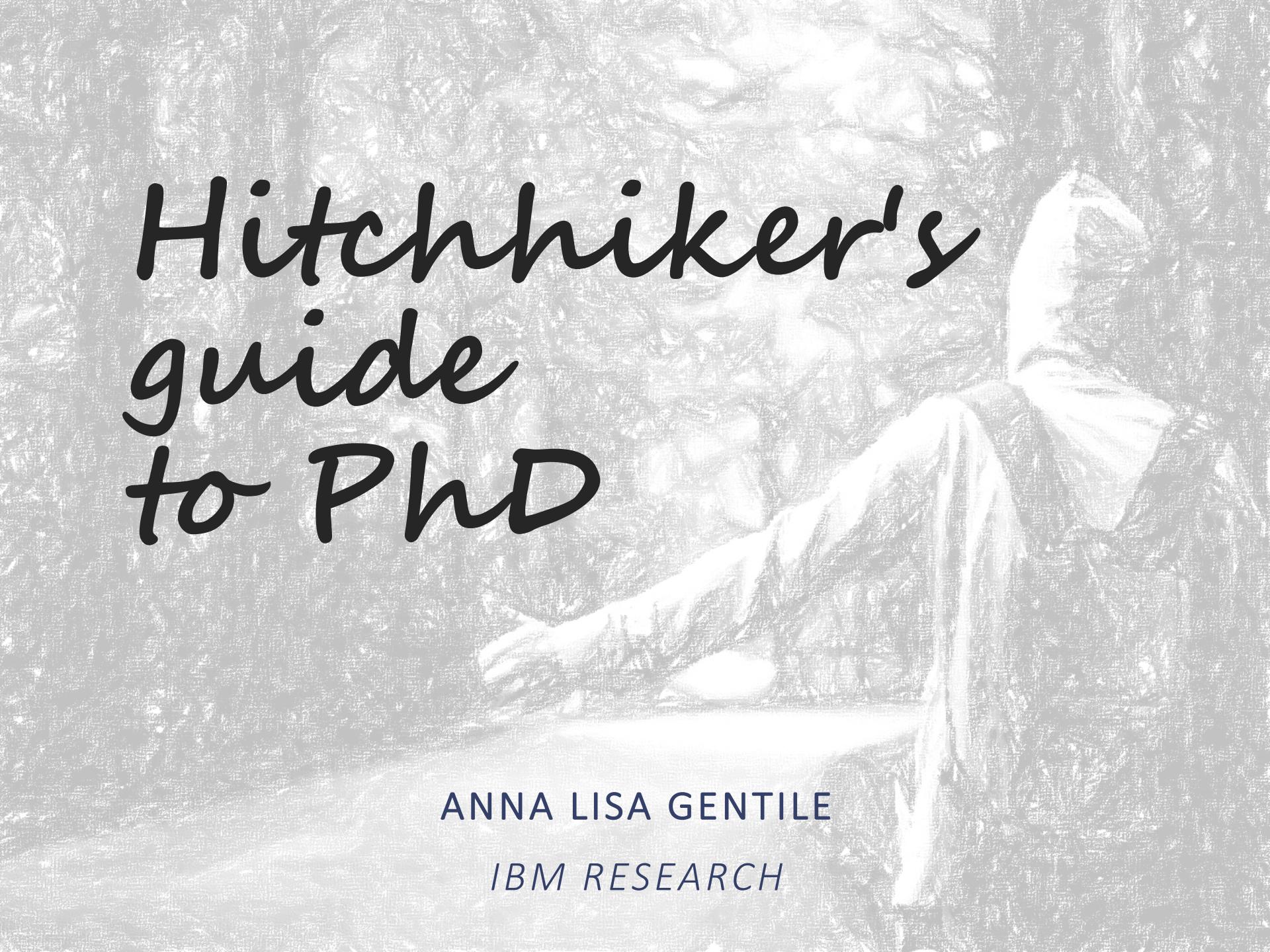


Hitchhiker's guide to PhD



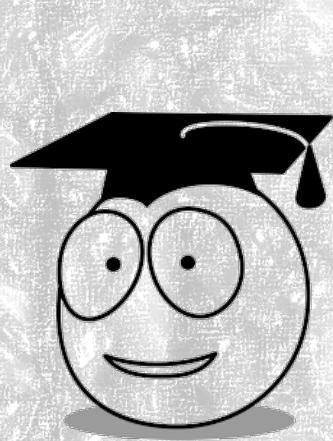
ANNA LISA GENTILE

IBM RESEARCH

About me



The
University
Of
Sheffield.



CALIFORNIA

What is this talk about

- How to do (Semantic Web) science
- Tips and tricks (from survivors)
- What happens after?

Credits

- How to do (Semantic Web) science
 - *Thanks to* Natasha Noy and Abraham Bernstein
 - *Primer* for Semantic Webbers [1]
 - slides based on [1] are marked with 
- Tips and tricks (from survivors)
 - *Thanks to* lovely SW friends for answering my *questionnaire*
 - slides based on their answers marked with 
- What happens after?
 - *Thanks to* the number 42



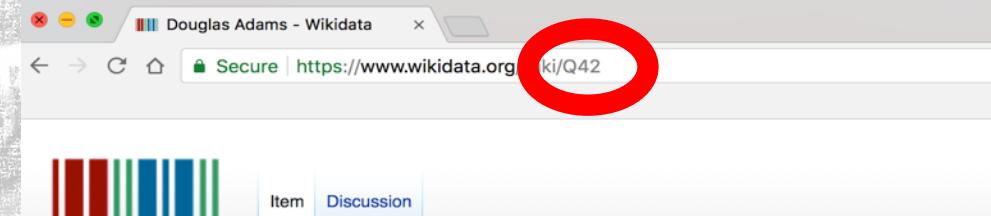
[1] Abraham Bernstein and Natasha Noy.

Is this really science? The Semantic Webber's guide to evaluating research contributions.

<https://www.merlin.uzh.ch/contributionDocument/download/6915>

The number 42

- the ultimate **answer** to life, the universe, and everything
- determined by *Deep Thought*, a supercomputer
- an even greater computer, *Earth*, is tasked to calculate the question to which 42 is the answer



The screenshot shows a web browser window with the Wikidata logo at the top. The URL in the address bar is <https://www.wikidata.org/wiki/Q42>. The page content is about Douglas Adams, with his name and the identifier Q42. Below the main content, there's a table showing Wikidata items for him, with English being the primary language.

Language	Label	Description
English	Douglas Adams	author and humorist
Italian	Douglas Adams	scrittore inglese

Semantic Web

- Extremely diverse research area
- Plethora of different Research Questions and methods
- What is a valid contribution?



Scientific Contribution

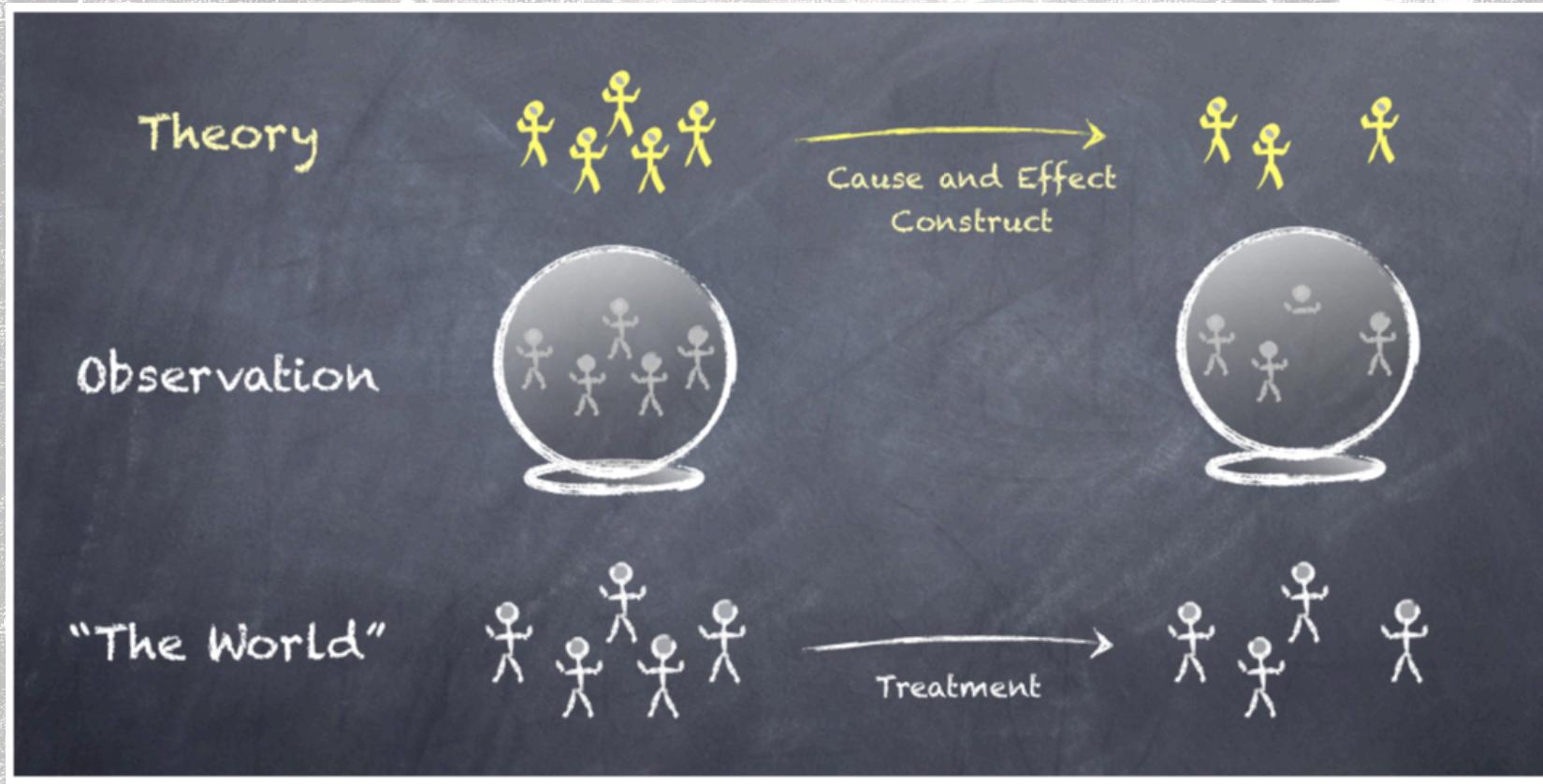
- Adding understanding on how something works in the *real world*
- What is the *real world* in SW?

“All science is either Physics or Stamp collection”

Ernest Rutherford



Physics and Stamp Collection



Physics and Stamp Collection

- gather observations about *the World*
- organize and structure them

example

International Classification of Diseases (ICD)
by The World Health Organization [2]

[2] <http://www.who.int/classifications/icd/revision/en/index.html>

[1] <https://www.merlin.uzh.ch/contributionDocument/download/6915>



Science: The Procedure

- Define **research questions**
- Formulate a **hypothesis** (or several)
- Design **evaluation** to test the hypothesis
- **Run the study**
- **Analyze** the results
- Report and **publish**



Research Questions and Methods

Does _____ (name the theory) explain the relationship between _____ (independent variable) and _____ (dependent variable), controlling for the effects of _____ (control variable)?

Creswell's (2009)

- Feasibility study
- Case study
(aka Demonstrator)
- Comparative study
(Benchmark)
- Observational Study
(a.k.a. Ethnography)
- Experiment
- Literature survey
(incl. Meta-Analysis)
- Formal Model
- Simulation



Validity and Reliability

- Construct validity
 - Is the way we are measuring misleading?
- Internal validity
 - Are the effects that we are measuring the result of using our method?
- External validity
 - Can we generalize the findings?
- Reliability
 - How consistent are the results?



Practical examples of Hypothesis: my System is better than yours

My reasoner is very fast and efficient **Bad**

My reasoner can classify SNOMED CT in 3 minutes **Ugly**

I can achieve a precision of 80% on entity extraction from a corpus X

My reasoner is faster than another reasoner X on a class of ontologies Y **Good**

Using X will improve the efficiency of reasoning on the class of languages Y , compared to the state of the art. **Excellent**



Practical examples of Hypothesis : the System

I have developed a system. It works!

Ugly

Domain experts can use our system effectively to
accomplish a task X (e.g., map between large
ontologies)

Better

Domain experts can use our system more effectively
than another system Z to accomplish a task X (e.g.,
map between large ontologies)

Excellent



Practical examples of Hypothesis : the world is better “in RDF”

I can convert this corpus of abstracts into RDF **Bad**

My conversion process produces better linked data than conversion
process X **Ugly**

Using extracted linked data will improve search performance on the
corpus compared to existing methods. **Good**

Using extracted linked data will enable advanced querying that was not
possible before **Good**

My method for extracting structured data has better accuracy/ cover-
age/precision/recall/etc. than the state of the art. **Good**



Tips & Tricks

I sent this **questionnaire** to some SW friends who survived their PhD in SW:

- What did you like
(things you would do again)
- What you did not like
(you wished you had done differently)
- Why did you choose Semantic Web?
- If it wasn't Semantic Web what would you choose?
- What advice(s) would you give to a PhD student

Tips & Tricks

Aggregated their answers as:

- Do
- Don't
- Why Semantic Web?
- If it wasn't Semantic Web then what?
- Advices



From SW PhD Survivors: Do

Link star
5-star
sciences
Internships
correlations
people
high-quality
publish
technologies
international
follows
ontological
Short background
want systems
formal
Especially
Description
Get better
always
Open
deeper
good
logging

IMHO XIX
Answer
problem DBpedia
course info
environment info
dereferencing
underlying
Serendipitously
URI
theoretical
pursue
disciples
enables
recommendations
committed
institutions
ending
community Great
exist
authors
information travel
computer web
social
inendisciplinary
LOD
generalized logics
and/or point
find
scientist's semantic makes
Collaborations biology
recommender work
comprises Engaged
allows predictions random
focusing Traversal actually turns
combine crawling already part
solving Logic



From SW PhD Survivors: Do

Work in an international environment

Collaborate across institutions

Travel a lot

Do internships

Great people in the community, make friends

Engaged community, make the most of it

Relations

Get a good understanding of the formal background (logics) in SW

Take advantage of Open data

Serendipitously crawl the Linked Data Cloud

SW specific

Interdisciplinary

“Solve existing problems using Semantic Web”

“Use Semantic Web for interdisciplinary research”

“Combine Linked Data with other disciplines”

Do high-quality research



From SW PhD Survivors: Don't

idea
didn't
argues
squeeze
favor
cleaning
already
Vibidato
shift
papers
sometimes
something
balance
consumming
long
cleaner
Hence work-life
Datasets

A word cloud centered around the term "data". The words are arranged in a roughly circular pattern, with "data" at the center. Other prominent words include "second", "community", "like", "industry", "using", "simply", "messy", "represent", "appropriate", "aspects", "Something", "hearing", "declares", "stifles", "tackling", "makes", "easy", "larger", "years", "state", "reuse", "attention", "representation", "concerned", "done", "syntax", "leave", "else", "timers", "even", "innovation", "others", "nothing", "aspect", "second", "done", "keep", "share", "just", "organize", "application", "self", "unless", "aspect...and", "still", "differently", "real-world", "code", "working", "needs", "part", "work", "design", "guard", "publicly", "poor", "ontology/vocabulary", "debating", "constant problem", "old", "SW-based", "right", "DBpedia", "Finding", "many", "benefit", "wasted", "put PhD", "focus last", "proper", "Gate hand", "right", "nigh

From SW PhD Survivors: Don't

- poor work-life balance
- poor self organization (this goes beyond PhD)
- messy code that you cannot share publicly
- don't ignore industry needs
- constant shift between aspects "correct representation of SW-based data" and "using LOD in its messy state"
- don't prioritize making data perfect, focus on tackling the problem of consuming messy data
- don't just squeeze the right keywords into your papers



From SW PhD Survivors: Why Semantic Web?

The SW
chose me ☺

It was a new and **exciting** technology at the time
- it is still exciting :)

The principle of representing knowledge **seemed fun**
- and thankfully still is

A **nice application** for logics

The romantic:

"I liked the idea of Linked Open Data
as well as the underlying SW technologies
and how it brings transparency"



The honest:

"because I started in a SW group and kinda liked it"



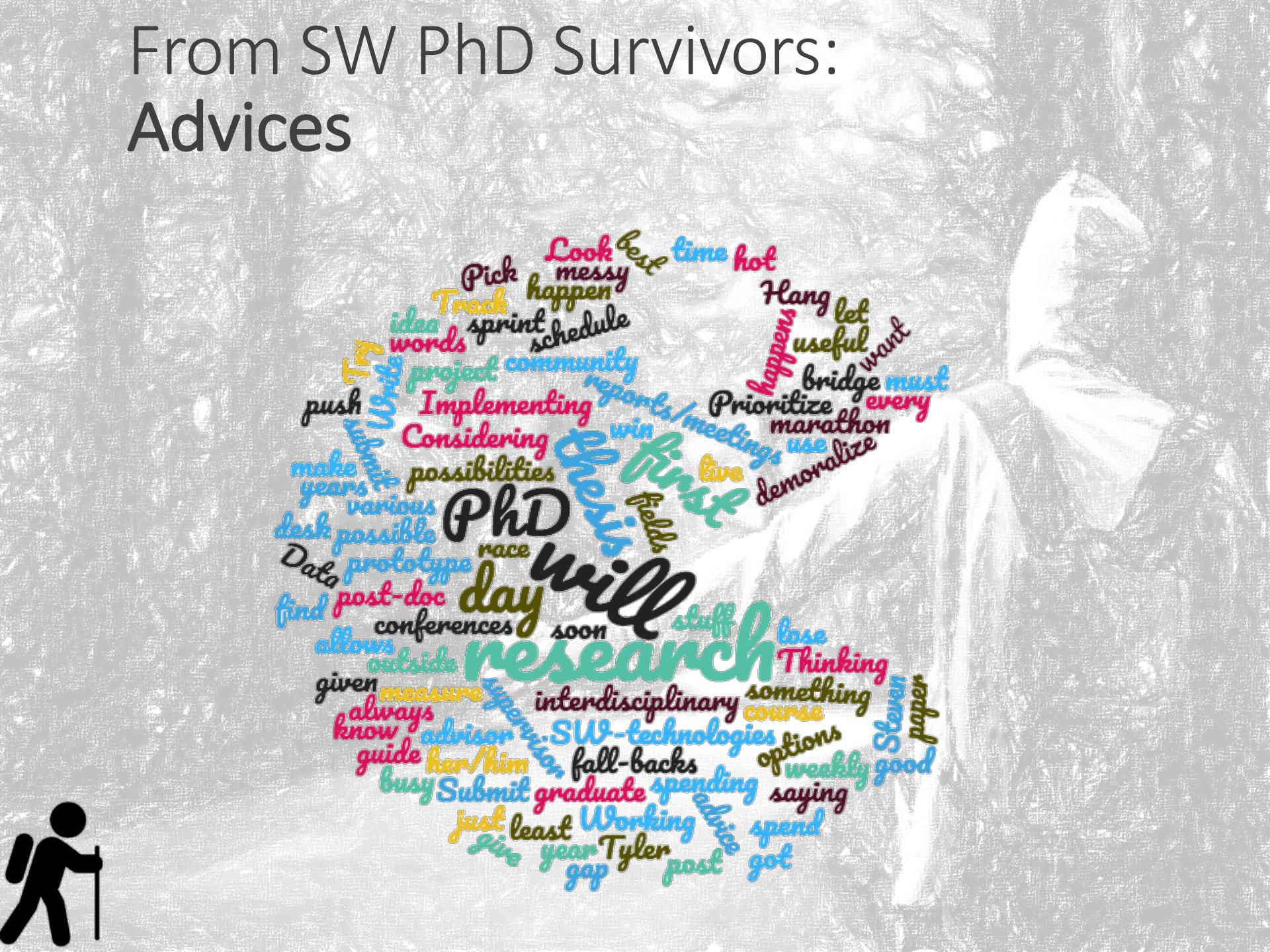
"You may say I am a dreamer, but I'm **not the OWLy one**"

From SW PhD Survivors: If not Semantic Web, then what?

Plain Machine-Learning
good Multi-agents
autonomous Neuroscience
old-data NLP
services science



From SW PhD Survivors: Advices



A word cloud composed of various advice terms related to PhD survival and research. The words are in different colors and sizes, centered around the word 'PhD'.

Key words include: Look best time hot, Pick messy, Track happen, idea sprint schedule, Try project community, push reports/meetings, Implementing bridge must, Considering win, Prioritize every, make possibilities, various live, Data prototype race, desk possible, will, find post-doc day, allows conferences soon stuff, lose, outside Thinking, given measure, interdisciplinary something, always course, know advisor, S.W.-technologies options, Stever paper, guide her/him fall-backs, weekly good, busy Submit graduate spending saying, just least Working, spend, give year gap, Tyler, post got.



From SW PhD Survivors: Advices

- PhD is a marathon, not a sprint race
- Hang in there: fall-backs happen as they must!
"You got to lose to know how to win" (cit. Steven Tyler)
- Find a **GUIDE** (supervisor, your favorite post-doc,...) as soon as possible



High level advices

From SW PhD Survivors: Advices

- schedule weekly reports/meetings with your supervisor
- Track what you spend your time in:
 1. Thinking about your thesis
 2. Implementing your thesis
 3. Working on other stuff
- Pick up 2 conferences where you want to submit every year. Submit.
- Prioritize your research, not the project where you are in



House keeping advices

From SW PhD Survivors: Advices

- Make something useful, not just a prototype
- Data are messy, live with it!
- Use SW-technologies in interdisciplinary research
- Look outside of the community and bridge the gap with other fields



sw advices

Tips & Tricks: Make the most of conferences

- Going to **talks** is not the only thing
 - networking, conversations, dinners, etc.
- Don't be afraid to talk to **anyone**
 - everyone is friendly, introduce yourself
 - have your **research pitch** ready
- Connect with your **peers**
 - they will be your colleagues in 10 years

Tips & Tricks: Research Pitch

“So what do you do?”

- Say it in *42 seconds*
 - Focus on your goals
 - Loose the details
 - What excites you about the work
- Say it in *three minutes*

...
- Discuss it for longer over a few beers

What happens after?

Sorry to disappoint you,
life remains the same

DON'T
PANIC



What happens after?

- The PhD is not your holy grail
- OK... you will probably get out with a few more **answers** than “the number 42”
- BUT... You will end up with a lot more **questions**
- Get done with it!

What happens after?

- The PhD is **not the end**
 - I actually started doing more exciting things during my postdoc(s)
- PhD → {postdoc}⁺ → Professor is not the only way
 - There are many alternative paths
 - It is OK to change mind

Hitchhiker's guide to PhD



annalisa.gentile@ibm.com



@AnLiGentile