

Building Knowledge Graphs in Organizations

Do you have a strategy?

Panos Alexopoulos

Head of Ontology @ Textkernel

<http://www.panosalexopoulos.com>

p.alexopoulos@gmail.com

@PAlexop

MY STORY



In 2016 I got a mission

**“Good morning Mr.
Alexopoulos. Your mission,
should you decide to accept it
is to build a Knowledge Graph
for the Labour Market Domain.”**

—Textkernel



which I accepted as I had the perfect recipe



that often didn't quite work as expected



leading to an important lesson



Never confuse movement with action.
- Steve Jobs

To live is to keep moving.
- Jay Shetty

Change is not a destination,
it's a process.
- Jay Shetty

You may not be interested in strategy,
but strategy is interested in you.
- Alan Watts

You may not be interested in strategy,
but strategy is interested in you.
- Alan Watts

To live is to keep moving.
- Jay Shetty

Never confuse movement with action.
- Steve Jobs

Strategies are not strategy. They
become strategy when you take
action.
- Jay Shetty

I HAVE A STRATEGY (NO YOU DON'T)

Never confuse movement with action.
- Steve Jobs

The essence of strategy is
showing what not to do.
- Michael Porter

The essence of strategy is
showing what not to do.
- Michael Porter

To live is to keep moving.
- Jay Shetty

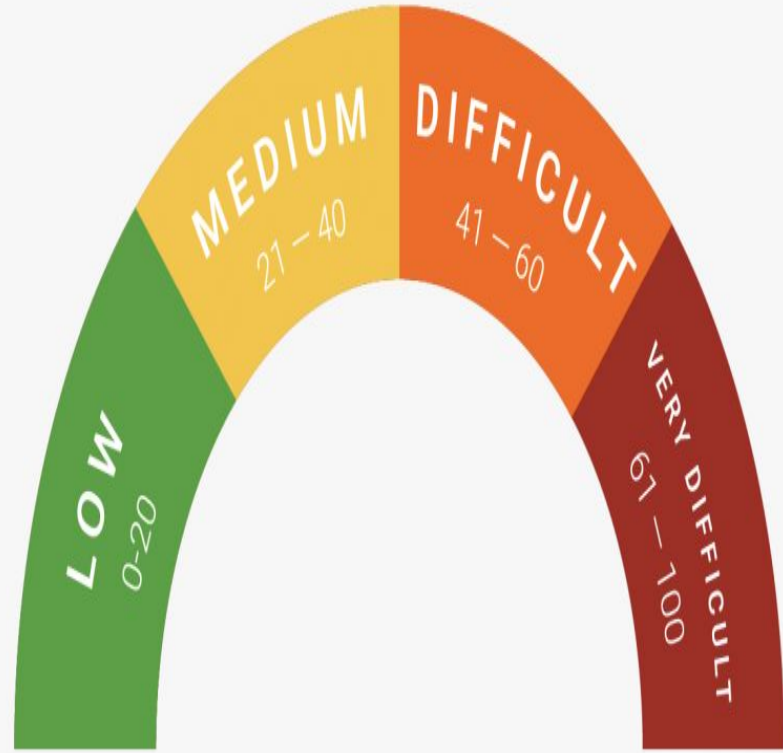
Change is not a destination,
just as hope is not a strategy.
- Ray Charles

Change is not a destination,
just as hope is not a strategy.
- Ray Charles



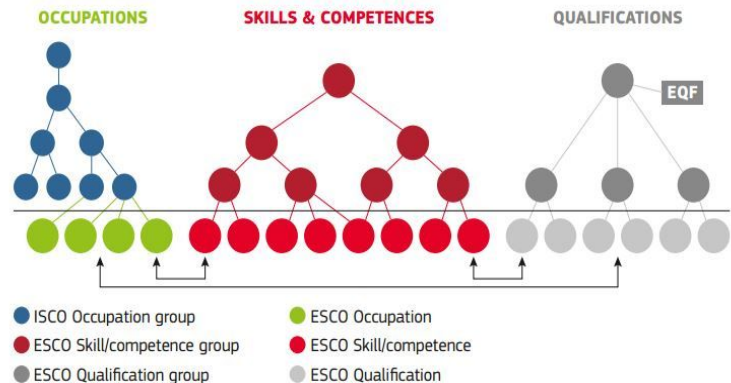
Building a Knowledge Graph

Is it easy or difficult?



A tale of two graphs

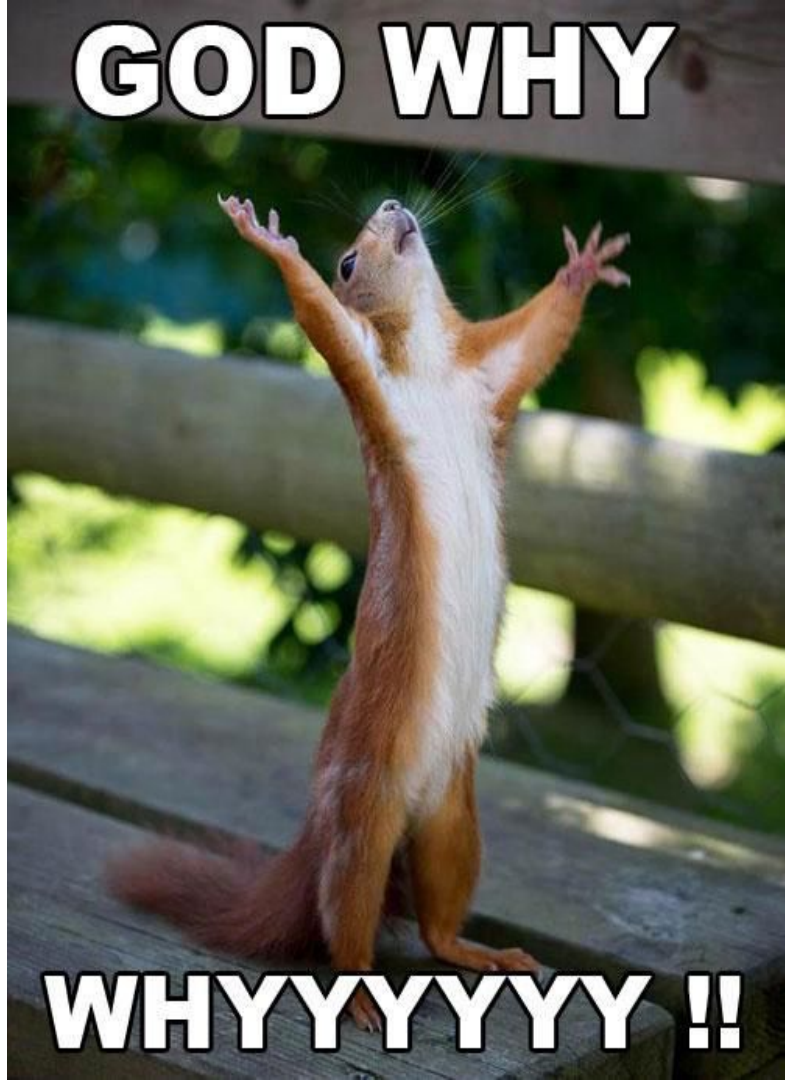
- **EU's ESCO knowledge graph:**
 - Released September 2017
 - 12K Skill and 3K Profession interrelated concepts in 27 languages
 - Built completely manually by experts in 6 years and cost ~15 million Euros
 - To be updated once a year by a maintenance committee
- **DBpedia knowledge graph**
 - First release in 2007
 - Currently >17M entities in 127 languages
 - Automatically extracted from Wikipedia infoboxes
 - Live version and monthly releases



Why the difference

- **Abstraction of entities:** Abstract entities more difficult to define than concrete ones (e.g. “Ontology” v.s. “The Netherlands”)
- **Vagueness of relations:** Vague relations harder to define and achieve consensus (e.g. “has essential skill” v.s. “was born in”)
- **Available data sources:** There are no Wikipedia-like infoboxes for Professions and Skills that could enable automatic extraction.
- **Quality requirements:** ESCO wanted very high accuracy and trustworthiness.

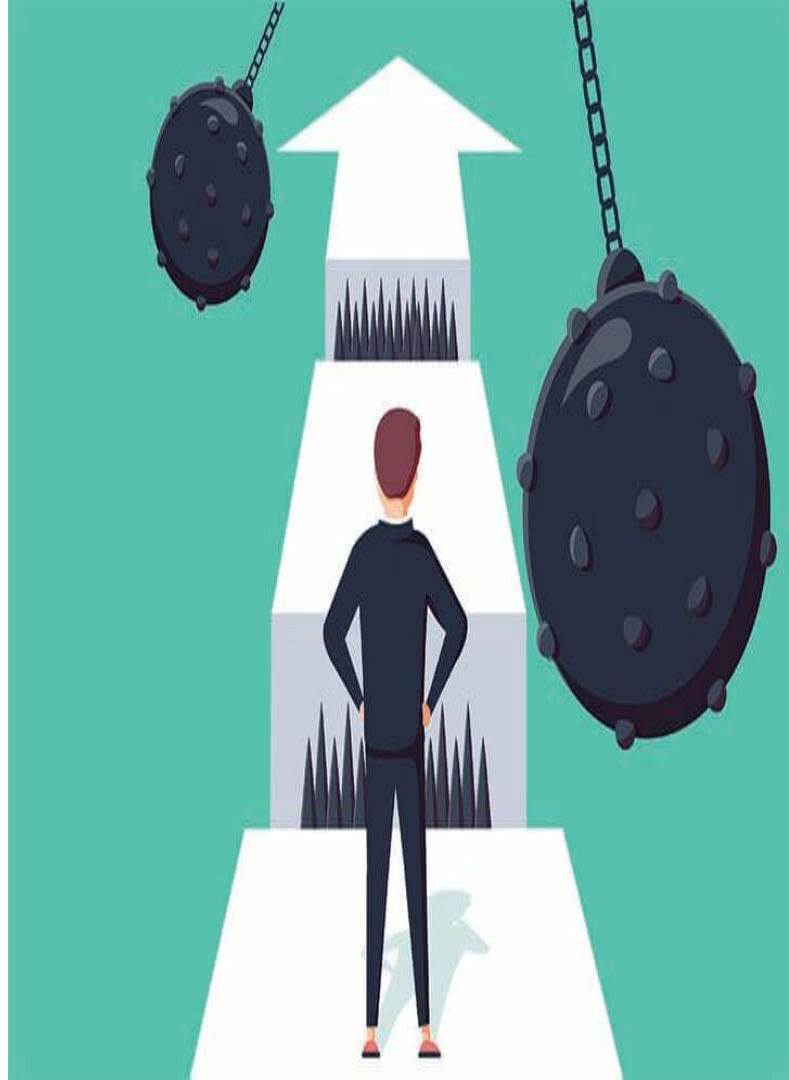
GOD WHY



WHYYYYYYYY !!

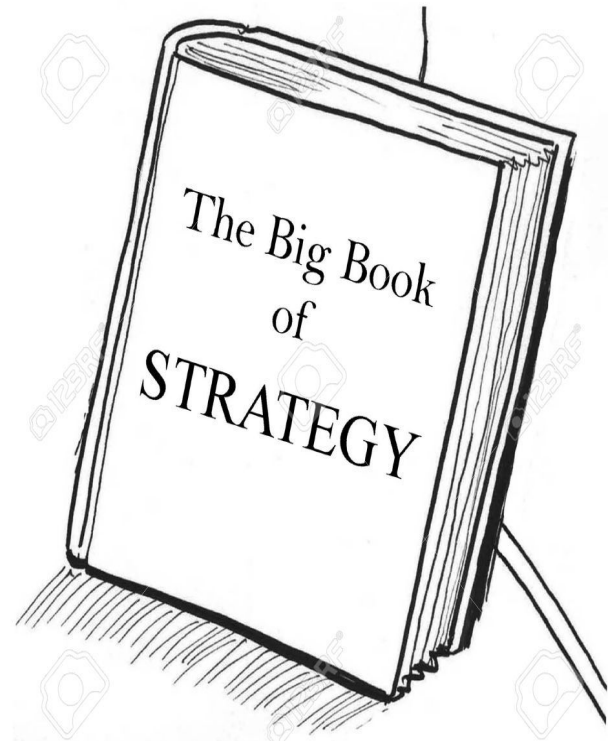
Difficulty factors

- **Clarity of semantics:** Abstract and vague entities/relations are harder to define/populate.
- **Mining/Extraction Data:** Availability, structure and appropriateness of data to automatically derive the graph from.
- **Maturity of Methods and Tools:** How effective and usable are algorithms and tools for automatic graph construction
- **Quality requirements/strategy:** Which quality aspects are more important and to what extent
- **Organizational Structure:** Existing data and knowledge management practices in the organization.



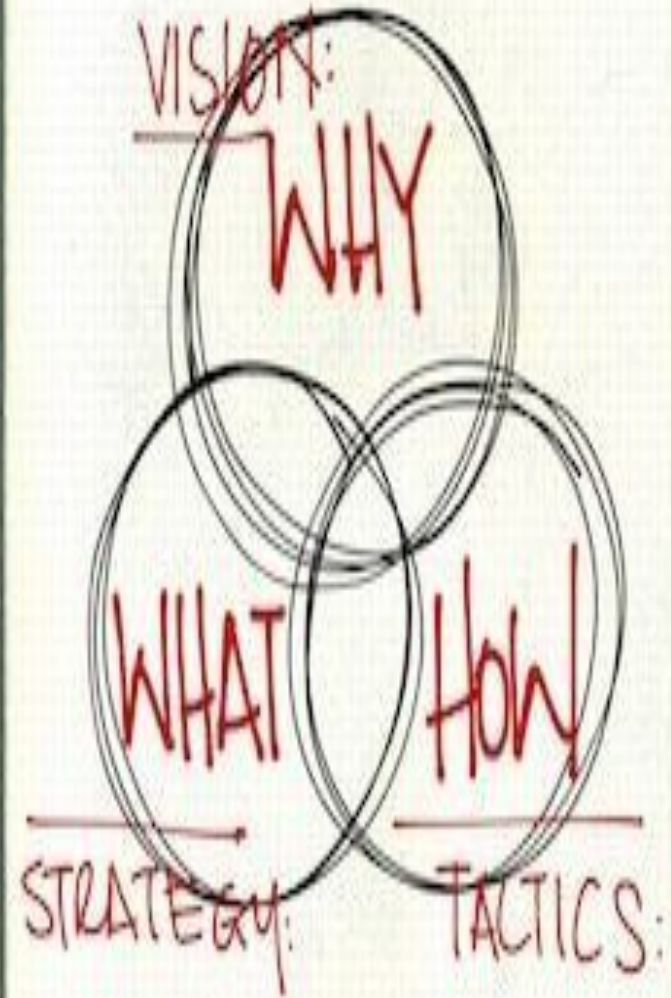
Knowledge Graph Strategy

What is it



Strategy is about “What”, not “How”

- **What** domains of knowledge we want our Knowledge Graph to cover?
- **What** data we want to connect?
- **What** applications we want to power?
- **What** users we want to serve?
- **What** level of quality we want to offer, have and in what dimensions?



Strategy is about Dilemmas and Choices

- Domain A or Domain B?
- Depth or Breadth?
- Accuracy or Completeness?
- Application A or Application B?
- Data-Driven or Human-Driven?
- Democracy, Oligarchy or Dictatorship?
- Single or Multiple Truths?



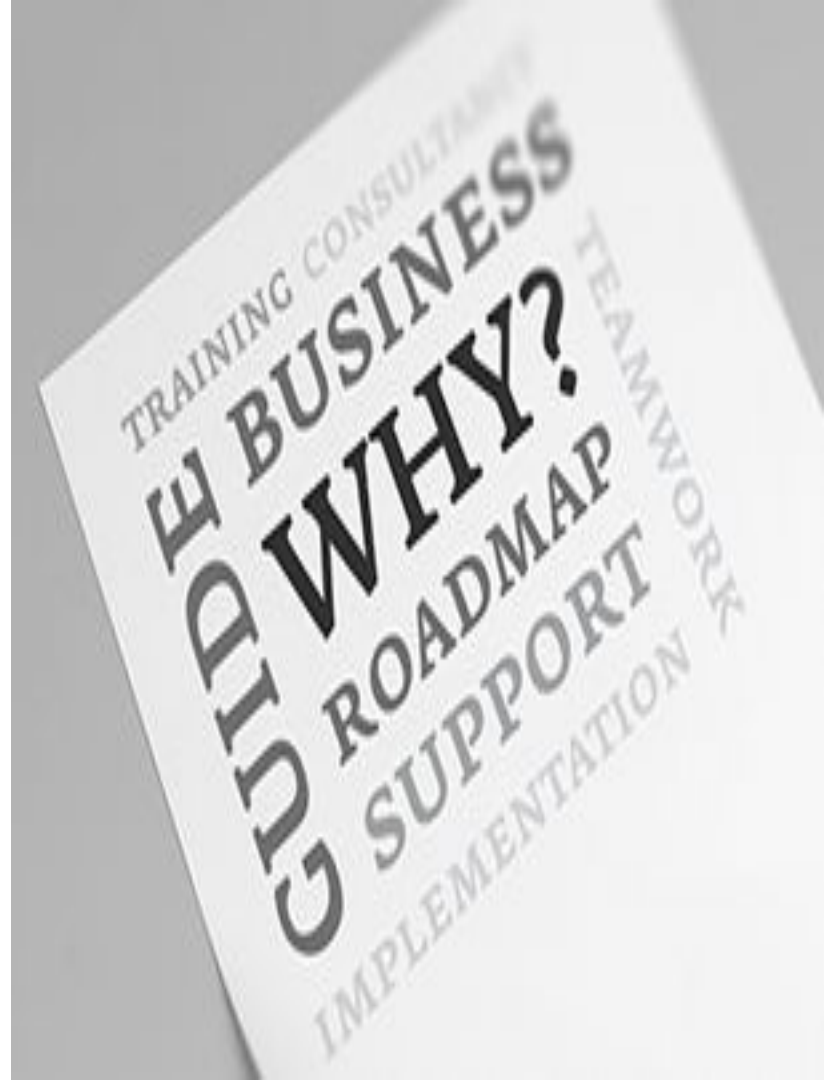
Knowledge Graph Strategy

Crafting one



Know thy goals

- What business/technical goals are you trying to achieve via the Knowledge Graph?
 - Imitate the others?
 - Enhance current products/processes and get ahead of the competition ? If so, which?
 - Lay the foundations for new products/processes? If so, which?
 - Integrate dispersed data and improve their governance?
 - ...



Know thy context

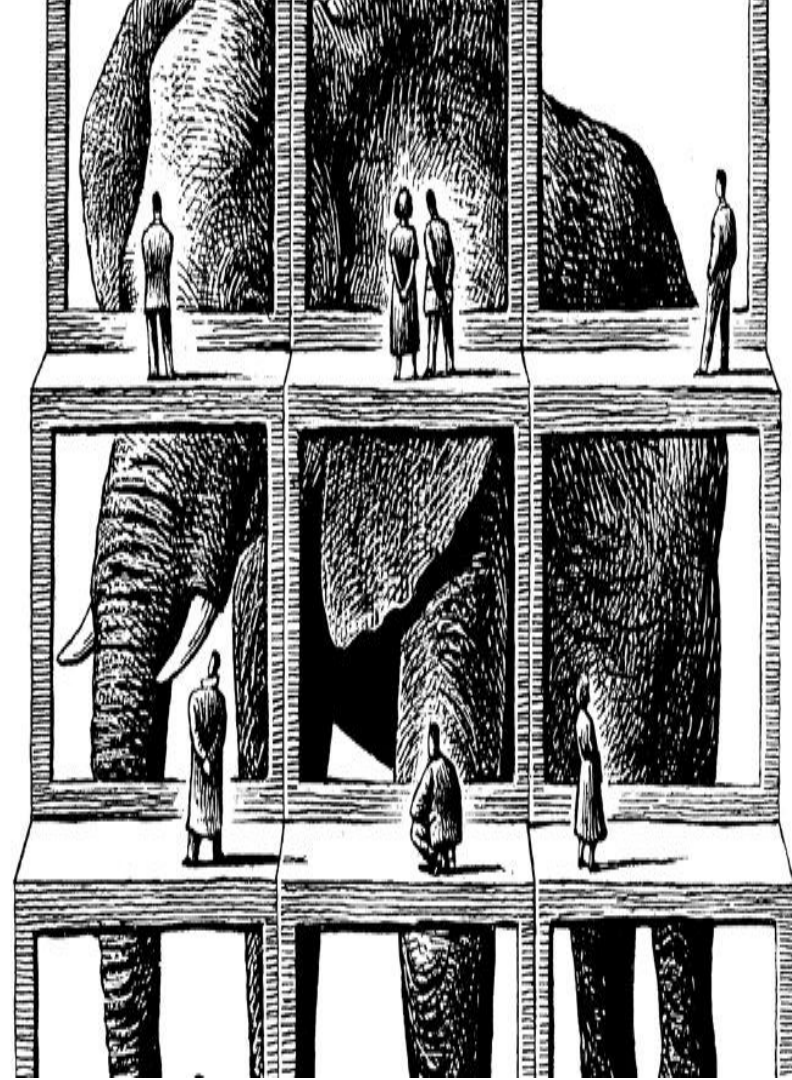
- **Domain(s) and Semantics**
 - What elements does the Knowledge Graph need to provide (entity types, relations, etc)?
 - Is there ambiguity, vagueness, semantic drift? To what extent?
- **Technology Context**
 - What knowledge can be automatically mined with the technology we have?
 - How fast and how accurately?

CONTEXT

MATTERS

Know thy context

- **Organizational Context**
 - Data, technology and products
 - People, knowledge and capabilities
 - Processes, practices and attitudes
- **Business Context**
 - Business strategy and model(s)
 - Clients, partners and competitors
 - Trends, hypes, expectations and narratives



Use context to make choices and break dilemmas

- If your semantics are highly vague and cause disagreements you might need to support multiple truths.
- If the best extraction system for relation X is only 50% accurate, you will probably need manual curation.
- If there is already a non-RDF semantics management infrastructure in the organization, it may not be worth replacing it with an RDF one!



Use context to make choices and break dilemmas

- If product A generates more revenue than product B, you might need to optimize the KG for A.
- If your competitor outperforms you in coverage you might want to compete on accuracy or expressiveness.





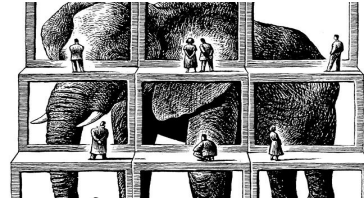
**KEEP
CALM
AND
WRAP IT
UP**

When you go to back to the office ask yourselves



Do we know why we are building (or want to build) a Knowledge Graph?

- High-level goals are ok but you need concrete ones
- Any answer with “AI”, “Big Data” or other buzzword should be ignored!



Do we know the complete context?

- Semantics
- Data
- Technology
- Organization
- Business

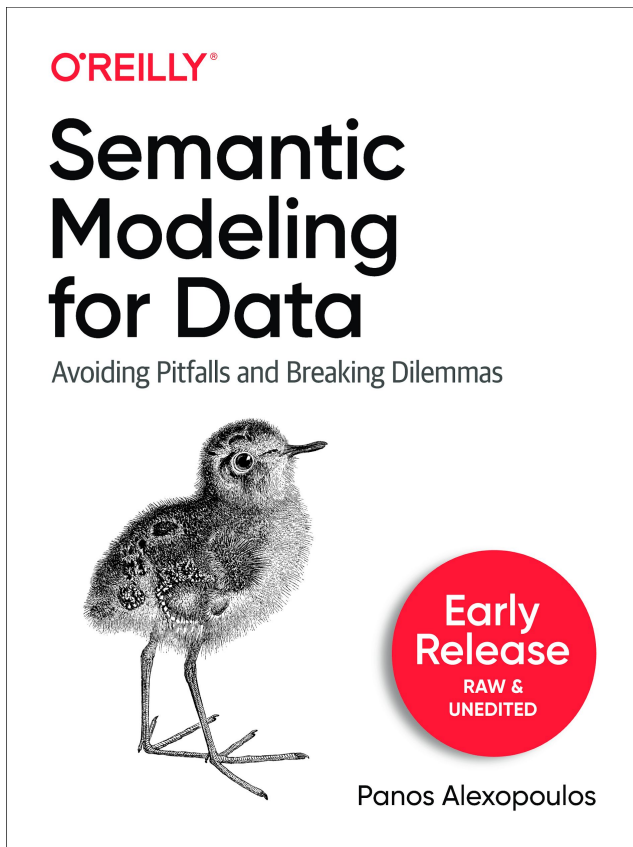


Do we know how to answer our “What” and break dilemmas?

- Basic philosophy and principles
- Definition of success (and failure)
- Decision making processes

For more knowledge graph pitfalls and dilemmas ...

- Scheduled to be published in September 2020 by O'Reilly Media.
- An Early Release version on O'Reilly's Learning Platform will be available next Monday.
- For more details talk to me or send me a message.





Thank you!

<http://www.panosalexopoulos.com>

p.alexopoulos@gmail.com

@PAlexop