2018-02-11

The story of my Ph.D.

The story of my Ph.D. is a story of bitter compromises. A story of compromising research for paid work. A story of abandoning promising research to be able to finish. Probably not that different than any other Ph.D.

Since memories are subjective and malleable, it may not be a completely accurate story. All departures from the true events are solely my faults. Moreover, I am not trying to convey any lessons learnt just yet. I just want to tell the story of *my* Ph.D. I guess it all makes *some* sense in retrospect, as if the future was a product of the past.

2010

While I could retrace ever so tenuous links way back, the story of my Ph.D. really began in November 2010 in Cologne, Germany. I was there for the Semantic Web in Bibliotheken (SWIB) conference as a co-author of one of the accepted papers. At the conference I met Sören Auer, there to promote the PUBLINK programme of the LOD2 project. While we talked over a coffee break I suggested that the Czech National Library of Technology, where I worked at the time, could use the consulting from linked open data experts offered in the PUBLINK programme. We exchanged business cards and parted our ways. It was the last time I had business cards. They were colourful and sloppy, manually cut from a thin paper. Still, they helped me to forge one of the most impactful connections for my Ph.D.

Weeks later I received an email from Sören, who concluded that we are qualified enough not to need help from the LOD2 project, and instead asked if we would consider joining the project as partners to represent the "Eastern Europe". Well, Czechs like to think they are a part of the "Central Europe", but I gladly seized this opportunity.

2011

After some discussion it became clear that the National Library of Technology did not have the workforce required to join the LOD2 project. I turned to my next closest institution: the University of Economics in Prague, where I already worked on bibliographic linked data that paid for my November trip to Cologne. There was a small team, led by Vojtěch Svátek, already involved in semantic web research for a number of years. To make this team stronger, we formed a strategic alliance with the group of Martin Nečaský from the Faculty of Mathematics and Physics at the Charles University in Prague, thereby transgressing the traditional organizational boundaries. This union later proved successful and lasted through many research projects we worked on together. It perhaps contributed to our affiliations blending in the minds of our foreign project partners to a nebulous concept of the "University of Prague".

Having secured a team, we needed a challenge it could work on. I started writing down a proposal that later turned into a part of the LOD2 project applying [linked open data for running a distributed marketplace of public sector contracts](). I based it on a suspicion that linked open data can serve as a better infrastructure for online markets. In such infrastructure, I surmised, we could operate matchmakers to link relevant demands and offers.

No idea is truly novel, and this one was no different. Its key inspiration came from [Michael Hausenblas](), whom I met at the Linked Data Research Centre at DERI (now [Insight Centre for Data Analytics]()) in Galway, Ireland, where I worked as an intern in 2010. Michael had similar thoughts earlier and came up with [Call for Anything](), a lightweight vocabulary for machine-readable descriptions of demands on the Web, and prototyped an application matching developers to businesses using the vocabulary. There already was a well-known vocabulary for describing offers on the Web: [GoodRelations]() by [Martin Hepp](). Call for Anything and GoodRelations clicked into place and I exchanged emails with Michael and Martin, thinking through an example application of matchmaking, which informed what later became our contribution to the LOD2 project.

What we lacked was a market in which data on both supply and demand is available. Earlier in 2010, when forming the nascent Czech open data initiative, we picked public contracts as a high-value dataset to screen-scrape and release as machine-readable linked open data. Public procurement market seemed to be a great setting for our work, since public contracts are demands explicitly represented as data as public procurement notices thanks to their proactive disclosure mandated by law.

We thought these poorly formed ideas through, enveloped them in profound academese, and eventually submitted them as an extension proposal for the LOD2 project.

The proposal was successful and in September 2011 we joined the LOD2 project, getting us three years of funding. It would be a perfect time for a Ph.D. if only I had already completed my Masters. I faced a dilemma later prominent in my Ph.D.: compromising paid work for educational progress. Moreover, back then I still worked part-time at the National Library of Technology. Nevertheless, I decided to fit everything in my limited waking hours and joined the University of Economics as an external researcher working on the LOD2 project.

2012

By the end of 2011 it was clear to me that splitting my time between work and education leads to hardly any progress in either of them. As my former position was no longer tenable, in February 2012 I quit the National Library of Technology to focus on finishing my master's thesis.

In the following months I started running up into the limits of part-time contracts at the University of Economics. The only reasonable way for me to work more on the LOD2 project was to enroll in the university's Ph.D. programme in applied computer science. There were no research ambitions at the beginning of my Ph.D. What made me apply for it was a practical concern to be able to continue in work that I found interesting. I applied for the Ph.D. and successfully completed the admission exams in May 2012. When in June 2012 I

graduated from the Charles University with a master's degree in [new media studies](#), I was set to pursue the Ph.D. However, even back then I was asking myself: [Is there life after Ph.D.?](#)

I officially began my Ph.D. on September 20, 2012. I started it believing the widespread myth that Ph.D. is the only opportunity in life to focus on a single thing and explore it in depth. I quickly realized how far detached from the reality this myth is.

Besides researching your Ph.D. topic many other duties compete for your attention. First and foremost, as a Ph.D. student I was required to teach. A cynical view has it that Ph.D. students are little more than a cheap resource to provision teaching. I was fortunate enough to be assigned with courses at least tangentially related to my Ph.D., including labs in an XML course and several lectures and labs in a course on linked data. The less lucky ones ended up teaching things like the basic Microsoft Office skills.

While teaching can be satisfying and meaningful at times, it also takes a huge amount of time to do it right, especially when you start a new course. The effort spent on teaching has sporadic returns. Rarely you hear any positive feedback, and given that one of the university's primary goals is to produce the most graduates, you often experience frustration with disinterested and unmotivated students who expect their graduation to be simply a matter of time. Under this impression, after a year, I decided to forfeit the Ph.D. stipend in order not to be required to teach.

Compared to teaching, other Ph.D. duties were relatively minor and infrequent. Once in a while I had to supervise bachelor's or master's theses and oversee admission exams of new students. I enjoyed the apprenticeship of supervising theses more than teaching, although few students invested more than required for a minimum viable thesis. Then there were academic duties that went without explicit acknowledgement, such as peer review, contributing to the bulk of unpaid labour that an obedient member of academia delivers.

The courses I was required to attend were largely irrelevant to the pursuit of my Ph.D. While I endured a course in IT management, I wondered why the courses on statistics or programming were left out of the curriculum. In retrospect, probably the most relevant was the introductory course on basic scientific methods, though it was definitely rudimentary.

Let's talk money. My Ph.D. stipend amounted to 5400 CZK per month, which was 216 EUR, or 75.8 % of the minimal net wage at the time in the Czech Republic. Back then it was roughly what you would pay for renting a room in a shared apartment in Prague. Since the stipend could not cover the cost of living, I had to find other sources of income, most notable ones being research projects, typically involving uncertain part-time and fixed-term work. I was decidedly a part of the Ph.D. [precariat](#), always compromising my research for paid work.

2013

The habit of following interesting work led me astray from my Ph.D. from time to time. For instance, between January 2013 and May 2014 I followed an opportunity to work with friends from new media studies at the Charles University on a project using semantic web technologies for the long tail of the job market. Also in January 2013 I achieved a minor

impact of my Ph.D. outside of academia. I was invited as a (charmingly called) "ad-hoc expert" to the European Commission's Public Sector Information group, where I talked on the dire present and the bright futures of public procurement data.

Contrary to my expectations, my actual contributions to the LOD2 project were rarely related to my Ph.D. More often than not I ended up doing the grunt work of data preparation or was swamped in the project admin and the ever-present "dissemination". LOD2 project also allowed me to take the inverse role of what I asked for in 2010 at SWIB. It was the National Library of Israel to which I served as a linked open data expert in the PUBLINK programme. As a result, when the LOD2 project successfully concluded in September 2014 most of my Ph.D. work was still left to be done.

2014

When the LOD2 project ended my future funding was unclear. By that time our proposal for a follow-up Horizon 2020 project called OpenBudgets.eu was rejected.

I used the gap in funding to do a Ph.D. internship at Politecnico di Bari, Italy, joining the research group of Tommaso di Noia between October and December 2014. It was an easy choice. When I surveyed the research literature on matchmaking (the topic of my Ph.D. thesis), I found many links pointing to Bari. In a fortunate turn of affairs, I managed to obtain my university's internal funding just in time for this internship. Working through a tight series of deadlines I completed my last required Ph.D. courses and re-enrolled as a full-time student in order to be eligible for the internship stipend. This internship was in fact the only period when I could be entirely dedicated to my Ph.D. It was essential in building the fundamental parts of what later became my thesis. I can heartily recommend going abroad for a few months to do such an internship.

2015

Immediately after my Bari gig, in January 2015, I followed with a one-month internship at the University of Göttingen, Germany, working with library data on old prints. Here again, I returned back to my roots in libraries. Also, I received a decent funding that sorted out my financial situation for another month and filled in some gaps from the previous period that the university's stipend failed to cover.

Since the research project funding at the University of Economics dried out, I arranged a part-time job for EEA from February to September 2015 working on the COMSODE project. There, I assumed a role of data janitor, tirelessly ETL-ing many government datasets. It gave me a novel perspective on the well-known setting of EU research projects. Working for a commercial project partner meant two things improved significantly: management and funding.

A peculiar turn of events took place in spring 2015. While it previously came short, the OpenBudgets.eu project was eventually funded and we were expected to start working on it as soon as possible, despite any plans we made in the meantime. I reluctantly accepted a part-time involvement on the project, starting in May 2015. With mixed feelings, I asked for a

break from my Ph.D., lasting till September 2015 when my contract with EEA ended. Due to the workload I imposed on myself, I was simply unable to fit the Ph.D. in.

In order to maintain my sanity during my long Ph.D. journey I occasionally worked on things whimsical. One of these "extra-curricular" efforts was DB-quiz, a Wikipedia-based knowledge game imitating a well-known Czech TV show. I found these activities fulfilling, perhaps because they helped me establish a sense in my Ph.D. in opposition to a clear nonsense. Obviously, I could not settle for anything halfway, so I followed through with the joke to the very end and turned DB-quiz into an academic paper, later winning a prize for the best Ph.D. publication at the University of Economics.

2016

At the end of 2015 I found myself with barely any progress in my Ph.D. It started to dawn on me that, if I am to finish at all, I need to live off my savings for a while instead of always hunting piecemeal income. Consequently, since 2016 I started carefully reducing paid work to make room for research. Since then my savings followed a decidedly declining slope.

I dedicated most of January to preparation for the doctoral state exam required after 3 years of Ph.D. The next month I passed the exam, albeit with a barely satisfying performance, and ticked off another Ph.D. duty: submitted a paper to my university's Ph.D. symposium. With these tasks out of the way there was only one thing left to do: my thesis.

September 2016 marked the start of the final year-long grind on my thesis. In fall of 2016 I thoroughly redone the entire data preparation, meticulously documenting its every step and improving my crude data processing tools on the way.

In December 2016, while entirely immersed in ETL of public procurement data, I realized that I forgot about the deadline for the preliminary thesis defense. By the end of the fourth year every Ph.D. student at my university is obliged to defend an 80% ready thesis. I started hastily piecing up my notes and former publications to meet my deadline coming up in February.

2017

My thesis had meagre 60 pages when I submitted it to the preliminary defense. Yet I managed to conditionally pass the defense thanks to otherwise outstanding results, judged by the modest standards of my university. The stipulated condition was that the thesis would be reviewed once more before the final defense.

Amidst the continued demise of my savings I was running out of options. I was piecing my income by context-switching between several part-time projects at my university. I needed another financial boost, one final kick before I was done with the Ph.D. The strategic alliance with Martin Nečaský came in handy again. Via this link I became a part-time open data expert at the Ministry of Interior of the Czech Republic, working on linked open data in statistics. During the summer of 2017 I pooled my time between this job, my thesis, and OpenBudgets.eu.

The problem with Ph.D. is that it [grows without bounds](). There is no way of telling that a Ph.D. is done, except the (arbitrary and untimely) end you set for yourself. When by the end of June I signed up as a full-time (linked) data engineer in pharmaceutical industry starting in October, I knew I had exactly 3 months left to finish my thesis. After that point I assumed there would simply be no time to work on the thesis anymore.

With a self-imposed deadline in sight I accepted the need for a bitter compromise. Compared with the original ambitions I left out many interesting experiments to be tried out. By the end of September I was mostly done, given my reduced work scope. Consequently, I passed the additional thesis pre-defense with no problems, giving me a green light to submit the final version of my thesis.

Unfortunately, despite my careful planning I did not manage to hand in my thesis before starting a full-time job. Due to an illness the thesis writing spilt by some weeks into October, with me working the evenings on final editing. I submitted my Ph.D. thesis on October 18. I was done. In relief, I [tweeted]():

> Submitting your Ph.D. thesis feels like a large open wound you've been bleeding from for years finally started healing. A nice feeling.

Apart from my turning in the thesis there were several other supplements I had to provide, the most puzzling one being a 20-page summary of the thesis. Frankly, who reads a 20-page summary? I thought people read either the abstract or the whole damn thing. Reservations aside, I bit the bullet once again and played by the rules.

THE END

On January 25, 2018, I passed my Ph.D. viva, with all committee votes unequivocally supporting my graduation. Finishing the Ph.D. was, first and foremost, a testament to my stubbornness, not to my research prowess. It took me 5 years, 4 months, and 5 days. It takes a lot of patience and grit to persist that long.

People are ridiculously bad at answering "what if" questions. Nevertheless, I believe I would not sit idly had I not enrolled in the Ph.D. I believe I would be doing something just as interesting. Hence, my overall evaluation of the Ph.D. is [exactly neutral]().

However, I could not disregard the Ph.D.'s negative externalities. The Ph.D. levied a toll on my relationships with others. Oftentimes I grew cold, moody, and unresponsive, as I was churning through the flexible working hours for a precarious income. I definitely was not the cheeriest lad around.

While I explicitly avoid any lessons for others here, there were lessons I learnt. I have grown ever so cynical. I have learnt not to care much for deadlines, adopting the attitude of Douglas Adams: *"I love deadlines. I like the whooshing sound they make as they fly by."* Finally, I have learnt to understand that *["adversity and existence are one and the same."]()*