

Summary of AAA Data Science Projects

AAA Researcher	Project Title	Project Status (July 2017)		
		A highlight in past 6 months	A highlight since start of project	A goal for next 6 months
Business Analytics				
Masoud Mazloom	Brand Centrality, Emotions and Virality in Social Media	Paper published in ACM ICMR, Bucharest, Romania, 2017: "A Spatio-Temporal Category Representation for Brand Popularity Prediction".	Proposing a multimodal approach for predicting the popularity of brand-related post/brand using visual and textual channels. Publishing several papers related to these topics.	Teaching a course on Fundamentals of Data Science and organising a data science summer school in Tehran, Iran 2018.
Qingchen Wang	Cross-Channel Multiple-Device Conversion Attribution	Two working papers: The first is in the final results stage (soon to be submitted for publication); the second has been submitted to the INFORMS Data Science Workshop.	Two working papers.	Joint project with a company, and to make a significant contribution to business academia and industry after data collection and optimized process is implemented.
Martijn van Otterlo	Data-Driven Decision Support in Complex Market Environments	Invited speaker at international library conference and distinguished paper award.	Finding a research problem, obtained funding for a public library project and developed first pilot project (indoor navigation in a library), published results in several papers. Developed new course on ethics of algorithms.	Second phase of the library project (more comprehensive implementation, live data) and working on algorithmic methods for behavior analysis of obtained data (and data from a related application).
Rens Dimmendaal	Innovation and Data Science: Generating knowledge on the success factors of mobile apps	Working paper on Optimal Distinctiveness of Mobile Apps was presented at the SAMS Creative Industries Paper Development Workshop.	The project started in Sep-2016.	Submit the first paper for conferences, and start the second paper in parallel.
Digital Humanities				
Serge ter Braake	Quality and Perspectives in Deep Data: Representation of Data Perspectives	Invited talk concept drift and digital methods at symposium Dutch Royal Library.	Interdisciplinary position paper on concept drift and digital methods.	Paper + use case concept drift of 'science', 'medical science' and 'vaccinations'.
Davide Ceolin	Reasoning on Information Quality Signals	Talks at DHBenelux and ADS controversy workshops on the assessment of the quality of Web documents; support to the organization of the Lorentz workshop on perspectives.	Paper on user studies on capturing, collecting analysing Web document quality assessments provided by humanities scholars presented at WebSci and EKAW conferences.	Large benchmark of the quality of Web documents.

Chantal van Son	Quality and Perspectives in Deep Data: From Text to Deep Data	Organisation of the interdisciplinary Lorentz Workshop on Language, People and Perspectives in Text in Leiden, involving researchers from Computational Linguistics, Humanities, Social Sciences and the Semantic Web. This was joint organisation with all three AAA-QuPiD projects.	Several papers on annotation projects; creation of corpus containing texts about vaccinations; expert annotations of micro-propositions on corpus of factuality; extensive literature research on focus; first pilot studies on crowd annotations for focus; reimplementation of existing focus detection system.	To submit two papers about: (1) crowd annotations for focus, (2) improved system for focus detection in text.
Informatics				
Julia Kiseleva	Citizen Data Science	Set up conversational search workshop at ICTIR 2017. Research on deep learning and reinforcement learning for information retrieval.	The project started in Feb-2017.	Submission on inverse reinforcement learning for search. Survey on conversational agents for search.
Benno Kruit	From Data to Semantics (and Back Again)	Submitted short paper to the International Conference on Information and Knowledge Management.	Extensive literature review and evaluation of current state-of-the-art systems.	Paper on extracting common knowledge about attribute changes from the web.
Peter Bloem	Information Content of Knowledge Graphs	Building on earlier work from the Data2Semantics project, we developed a method, based on Minimum Description Length, to scale motif analysis up to billions of graphs (code , pre-publication). We cooperated with researchers at the UvA AMLab and the ILLC translate graph convolutional neural networks to the domain of knowledge graphs (pre-publication).	We presented an analysis of IRI structure across the LOD cloud at ISWC 2016 .	Work in progress includes learning sparse transformations on knowledge graphs, a variational method for knowledge graph node embedding, and an extension of our motif method to knowledge graphs.
Bart Kamphorst	Data-driven Lifestyle Support through Smart Devices	Designed and implemented software for staging adaptive e-health interventions via Telegram. An abstract pertaining to this work was submitted to the ISRII conference.	Authored a Technical Note (PR-TN-2016/00553) on machine learning techniques suitable for integration into behavior change solutions.	Perform a proof-of-concept study to evaluate the newly developed platform and write a paper detailing the findings.

Life Sciences				
Ted Meeds	Explaining Disease using Big Data: How Valid is your Pathway	Invited talks at Microsoft Research Cambridge and the University of Montreal.	The development of a new method that removes cohort-level effects in a pan-cancer model of cancer genomics data. This allows our model do learn disease profiles that are general across cancer types.	To finalize a publication based on integration of multiple molecular sources (RNA, micro RNA, and methylation) on a pan-cancer dataset.
George Kanev	Integration of Phenotypic Drug Efficacy and Molecular Chemogenomics Data	Integration of databases with genetic, phenotypic and pathway data to predict effective drug combinations in the AstraZeneca-Sanger Drug Combination Prediction DREAM Challenge, and a top 5 position (out of 100 teams).	Establishment of the kinase-drug interaction databases KLIFS and PDEstriAn to systemically discover drug-target interactions using all available kinase and small-molecule structural and bioactivity data, resulting in 3 publications.	Experimental validation of the effectivity of computationally predicted multi-drug combinations and optimal alignment of drug-target profiles to cancer vulnerabilities to improve therapy efficacy for a number of representative cases.
Frans van der Kloet	Network-Based Data Integration	DSM asked our support for analyzing complex data (50k; contract research, valorization).	In collaboration with stakeholders (VUmc, NKI) a proper problem was defined. A new method is under development to solve this problem.	Frans van der Kloet has left the project as of April 1. The new postdoc (Dicle Hasdemir) starts Sept 1, 2017.
Social Sciences				
Bob van de Velde	Bias and Engagement in Political Social Media	Launch of ROBIN plugin for data collection of personalized page content, Papers at Collective Intelligence and SIGIR, bookchapter on storing data for social scientists.	Engaged in two collaboration projects (ILPS-Political communication; UvA-VU).	Write papers on visual bias, news bias and personalized communication.
General				
Eva Kenny	Project Management	Coordination of 14 Meetup events for the Amsterdam Data Science community on a range of topics: Responsible Data Science, Machine Learning, AI, Education, Startups.	Launch of a new Amsterdam Data Science website, launch of a new School of Data Science website including populating with 200+ courses and programmes.	Assist in developing International collaborations with Data Science Centres in Europe and China.