



NATIONAL RESEARCH  
UNIVERSITY

Geoinformatics Group  
<http://geolab.gis.land/>

# Research Directions of the Geoinformatics Group

Ramon Antonio Rodrigues Zalipynis  
[arodrigues@hse.ru](mailto:arodrigues@hse.ru)

# Outline

- **R.A. Rodrigues Zalipynis** – about myself
- **What is geoinformatics?**
- **Geospatial Data**
  - Vector data
  - Raster Data
  - Big Data: 6Vs
- **Related work**
- **Research Directions & Project topics**
  - 1st year (course work)
  - 2nd year (M.S. Thesis)
- **Q & A**

# Ramon Antonio Rodrigues Zalipynis

Associate Professor: [Faculty of Computer Science](#) / [School of Software Engineering](#)

*Ramon Antonio Rodrigues Zalipynis has been at HSE University since 2015.*

[Home](#)

[Teaching](#)

[Research and Publications](#)

[Projects](#)

[In HSE news](#)



**Ramon Antonio**

[arodrigues@hse.ru](mailto:arodrigues@hse.ru)

# Ramon Antonio Rodrigues Zalipynis



## VLDB 2018

The only talk with Russian Affiliation @ VLDB for the last 10 years (2009–2018) and the 1st from HSE

## SIGMOD 2019

The 7th paper @ SIGMOD with Russian affiliation for all the years of the SIGMOD conference (1975–2019, 45 years) and the 1st from HSE

## VLDB 2020

The 2nd paper @ VLDB with Russian affiliation for the last 10 years (2010–2020) and the 2nd from the HSE

**SIGMOD /  
AMSTERDAM  
PXDS  
2x19**

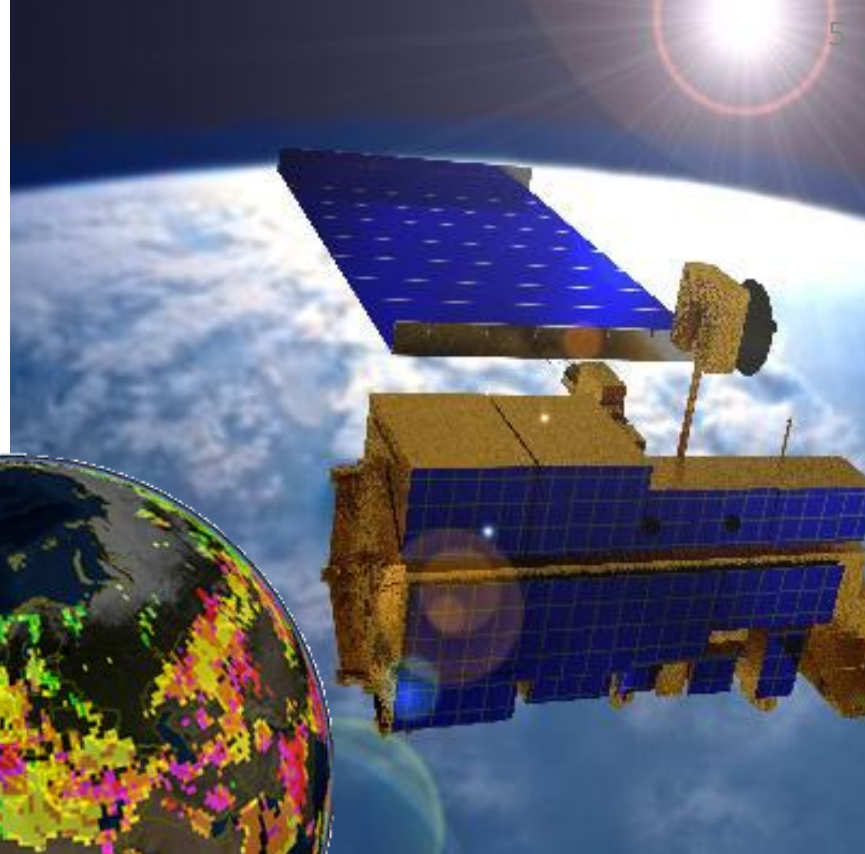


**VLDB & SIGMOD – top conferences on**

- Databases
- Big Data
- Distributed systems
- Data management

**Geoinformatics:** storage, processing, analysis and visualization of data with geographic reference

- Спутники, лидары, БПЛА
- Моделирование (погода, климат, гидрология, ...)
- Мобильные устройства
- Любые данные с геопривязкой



# Importance

## Big data

- ECMWF data warehouse: **137.5 M files, ~52,7 PB** (2015)
- DigitalGlobe, satellite data provider: **~80 TB/day** (2017)
- Planet: «Earth's selfie daily», **400 satellites** (2021)

## Commercial attractiveness

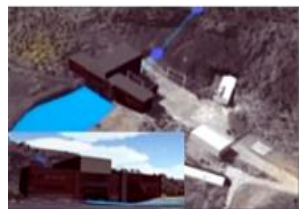
- Apple объявила сразу 70 вакансий «гео-специалистов»: <https://www.cnbc.com/2017/08/02/apple-has-over-70-map-tech-job-openings.html>
- Uber geospatial data visualization <https://eng.uber.com/keplergl/> 29 Мая 2018 г.
- Carto cooperates with MasterCard <https://carto.com/blog/carto-mastercard-partnering-location-intelligence-solution/> 22 Мая 2018 г.



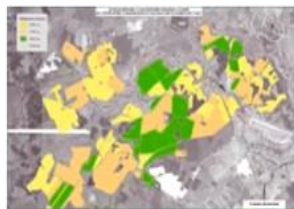
U B E R



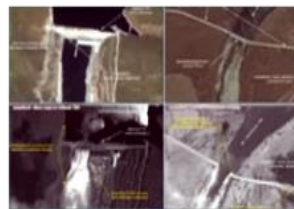
## Practical importance



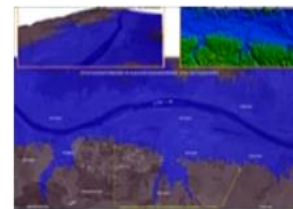
ТЕРРИТОРИАЛЬНОЕ  
ПЛАНИРОВАНИЕ



КОНТРОЛЬ  
ЗЕМЛЕПОЛЬЗОВАНИЯ



КОНТРОЛЬ  
ТЕХНОГЕННЫХ АВАРИЙ

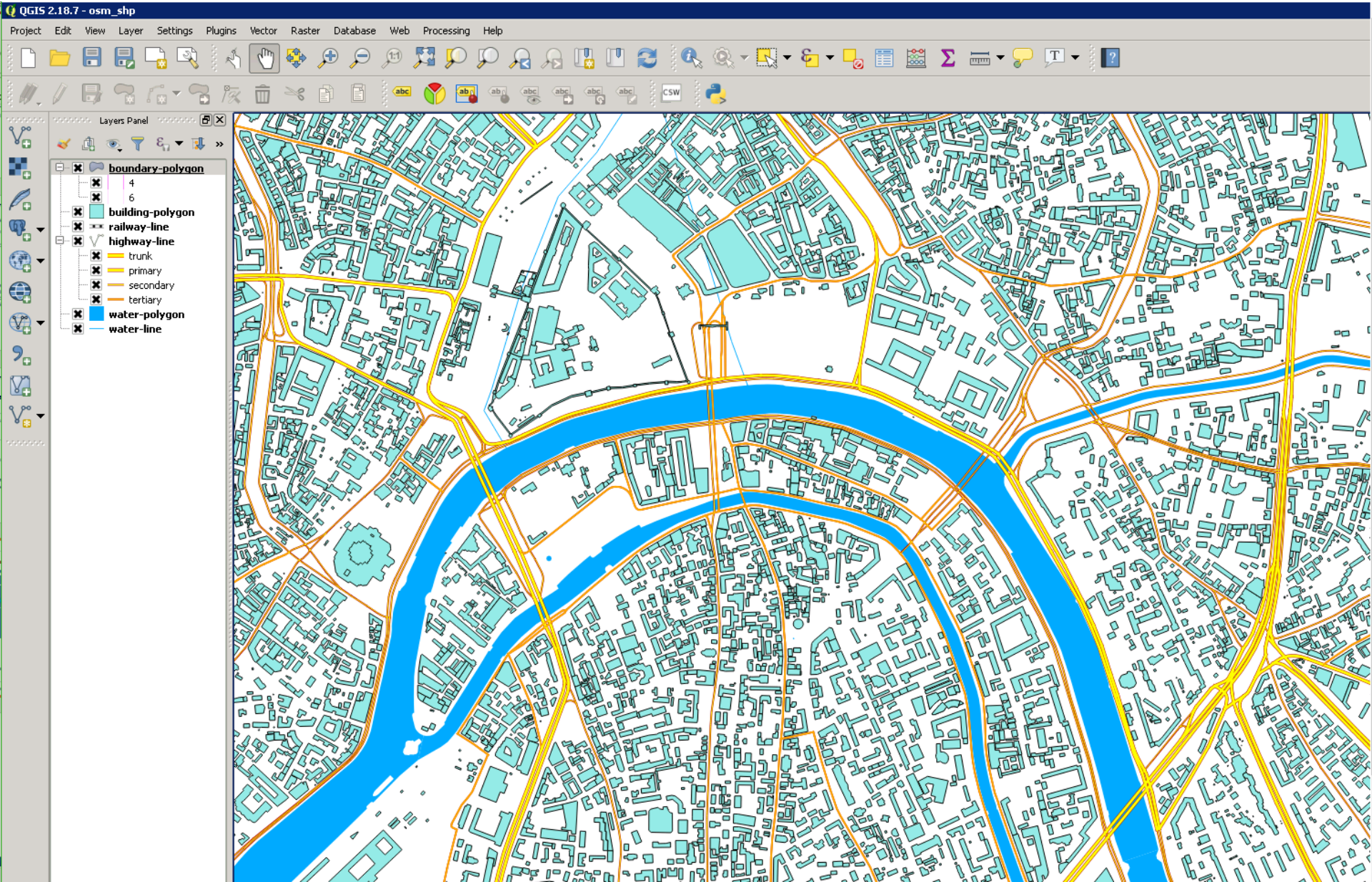


ОЦЕНКА  
ПАВОДКОВЫХ СИТУАЦИЙ



КОНТРОЛЬ  
ЛЕСНОГО ХОЗЯЙСТВА

# Vector data: Moscow



Raster data: 30 cm/pixel





# 6Vs

## Geospatial raster data – big data?

### VOLUME

The amount of data from myriad sources.



### VARIETY

The types of data: structured, semi-structured, unstructured.



### VELOCITY

The speed at which big data is generated.



### VERACITY

The degree to which big data can be trusted.



### VALUE

The business value of the data collected.



### VARIABILITY

The ways in which the big data can be used and formatted.



# Volume & Velocity: Maxar (former DigitalGlobe)

- **Collects 80 TB/day**
- **Accumulated in Amazon over 100 petabytes of satellite imagery**
- This is only one of the many companies and organizations

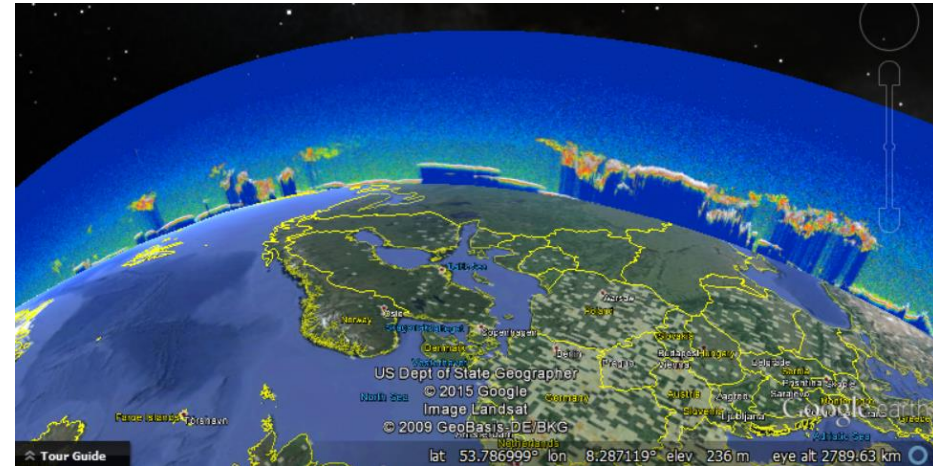
<https://youtu.be/mkKkSRixU8M>



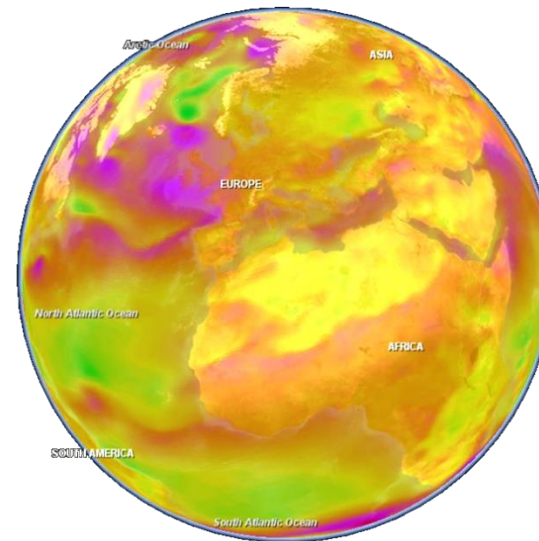
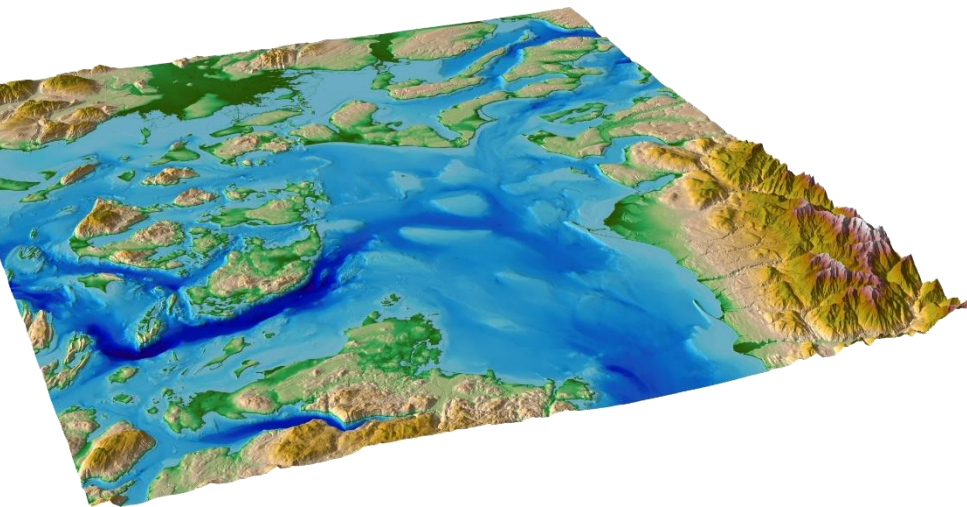
# Variety (разнообразие)

- 1D, 2D, 3D, 4D, ...
- 140+ formats
- Heterogenous structure

## Vertical aerosol profile (CALIPSO)



## Digital elevation model



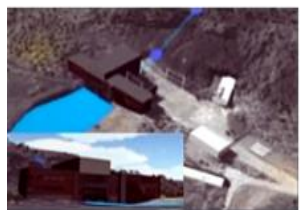
## Wind speed (assimilation, model)

# Value (польза) & variability (варианты использования) <sup>12</sup>

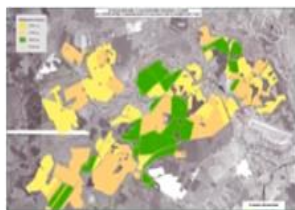
We can facilitate decision making in the following areas:

- **Чрезвычайные ситуации (emergencies)**: какие районы города будут затоплены при повышении уровня воды на 1 метр (используются DEM)?
- **Сельское хозяйство (точное земледелие, precision agriculture)**: как нужно удобрять поле, чтобы собрать больше урожая?
- **Городское планирование (urban planning)**: где и какие санитарно-гигиенические мероприятия нужно проводить, чтобы повысить качество воздуха?

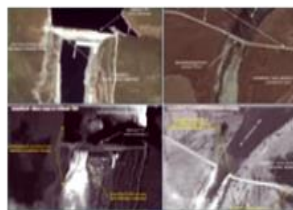
[https://www.youtube.com/watch?v=LsR\\_gfCeSIQ](https://www.youtube.com/watch?v=LsR_gfCeSIQ)



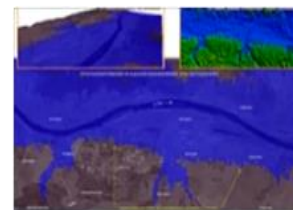
ТЕРРИТОРИАЛЬНОЕ  
ПЛАНИРОВАНИЕ



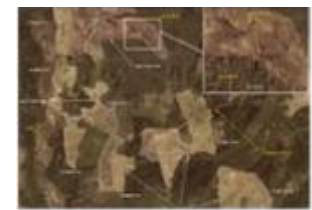
КОНТРОЛЬ  
ЗЕМЛЕПОЛЬЗОВАНИЯ



КОНТРОЛЬ  
ТЕХНОГЕННЫХ АВАРИЙ

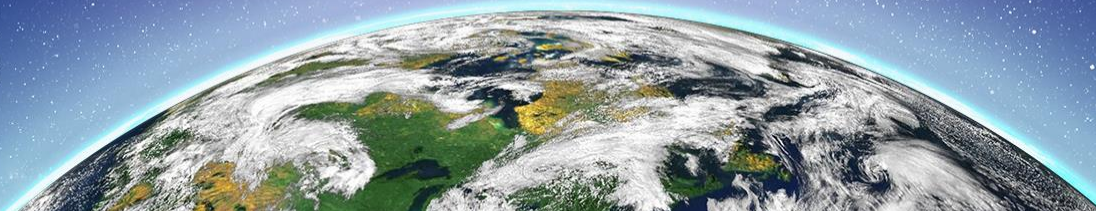


ОЦЕНКА  
ПАВОДКОВЫХ СИТУАЦИЙ









КОНТРОЛЬ  
ЛЕСНОГО ХОЗЯЙСТВА

<https://learn.arcgis.com/ru/arcgis-imagery-book/>



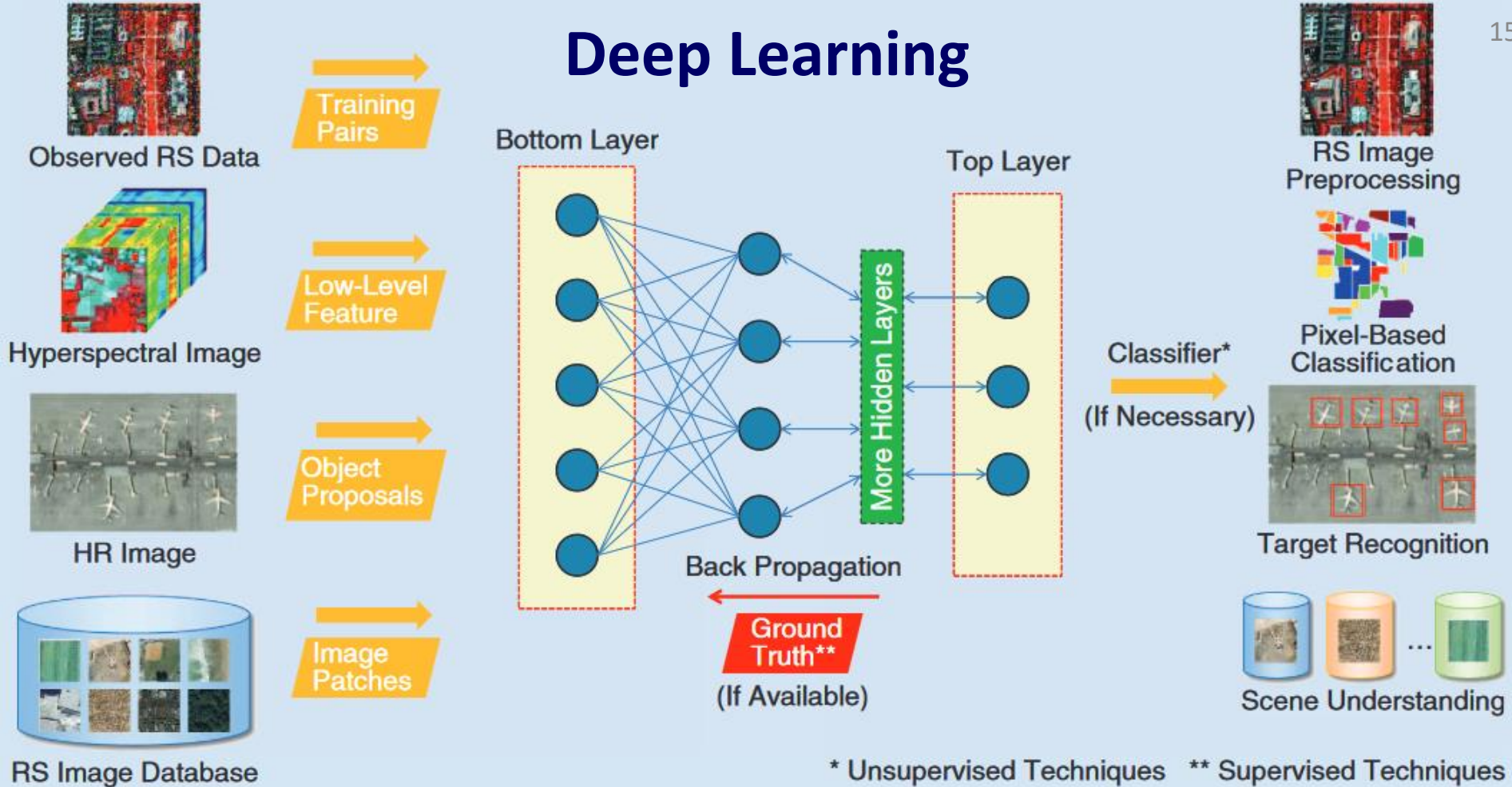
# Veracity – достоверность

- **We are dealing with physical measurements, not just RGB images**
- **Federal Law 2017 (лесной кодекс):** арендатор лесного участка обязан прикладывать к отчету об использовании леса материалы данных дистанционного зондирования Земли, подтверждающие соблюдение границ выделенного участка под вырубку
- **Forests in Russia – about 8,09 M km<sup>2</sup> (20% of all world forests)**

VOLUME	VARIETY	VELOCITY	VERACITY	VALUE	VARIABILITY
The amount of data from myriad sources.	The types of data: structured, semi-structured, unstructured.	The speed at which big data is generated.	The degree to which big data can be trusted.	The business value of the data collected.	The ways in which the big data can be used and formatted.
					

# Related work

# Deep Learning



- RS image preprocessing:** цель – улучшить качество входных данных (напр., убрать шум)
- Pixel-based classification:** классы – дороги, строения, тип сельскохозяйственной культуры, ...
- Target Recognition:** выделить и распознать объекты на сцене (напр., самолеты, корабли, ...)
- Scene Understanding:** что здесь расположено/что происходит? – чтобы оператор вручную не просматривал терабайты сцен высокого разрешения (карьер, аэропорт, вырубка леса, ...)

# Пример: оценка rooftop solar potential

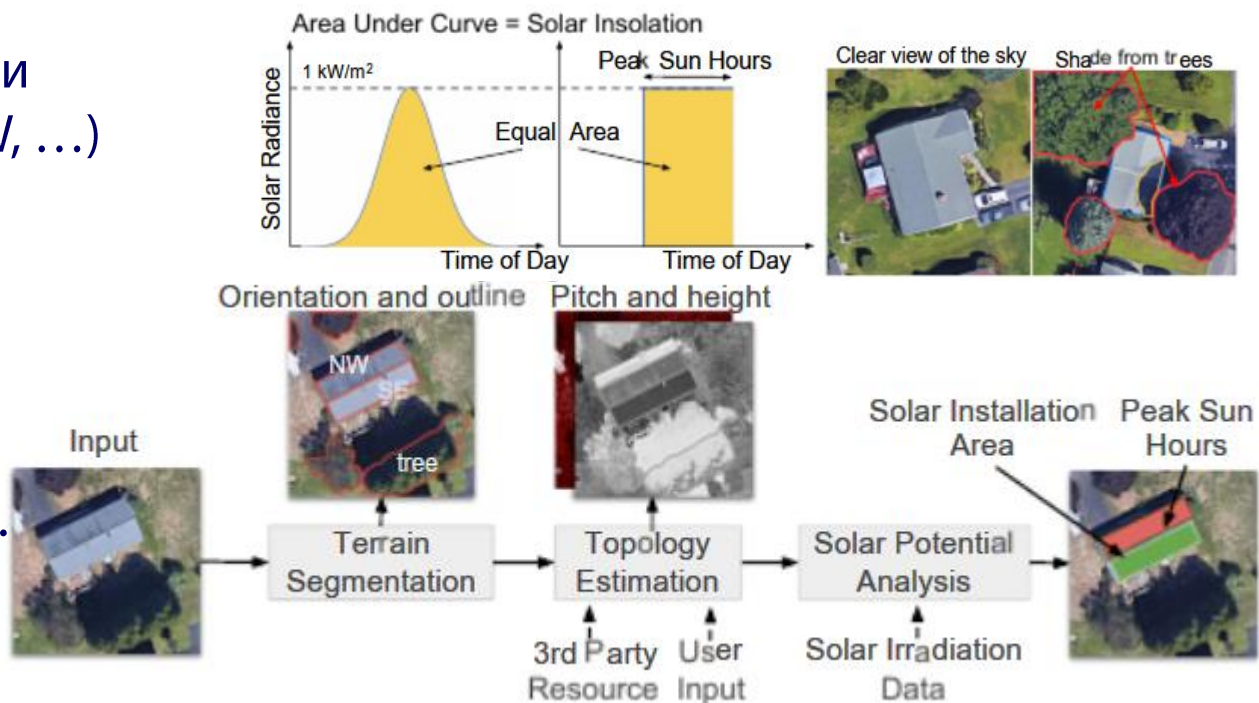


2016 год: новых батарей на 70 гигаватт [link](#)

- независимый источник green energy
- оценка потенциала – консультанты
- цель – массовая автоматическая оценка с помощью спутниковых данных

## Постановка задачи:

- для каждой крыши найти
  - ориентацию (W, S, NW, ...)
  - высоту
  - наклон
  - потенциальные тени
- разделить крышу на сегменты
- по формулам и вспомог. данным рассчитать потенциал

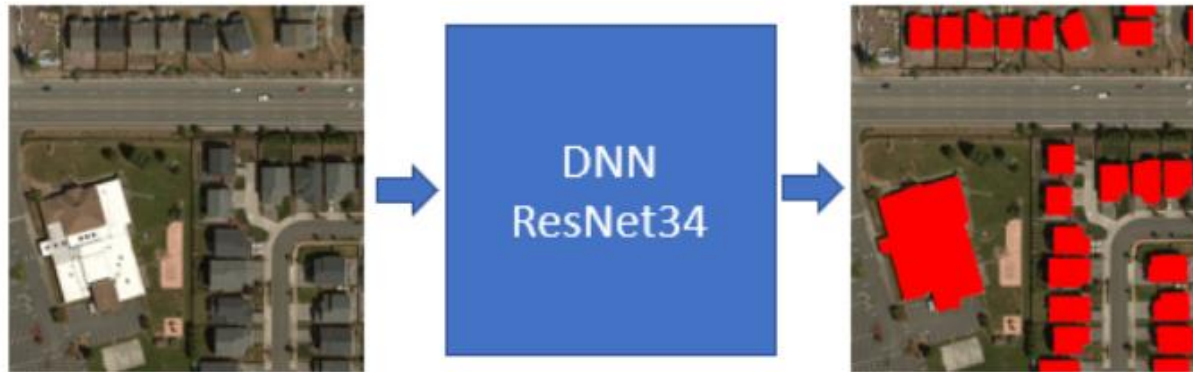




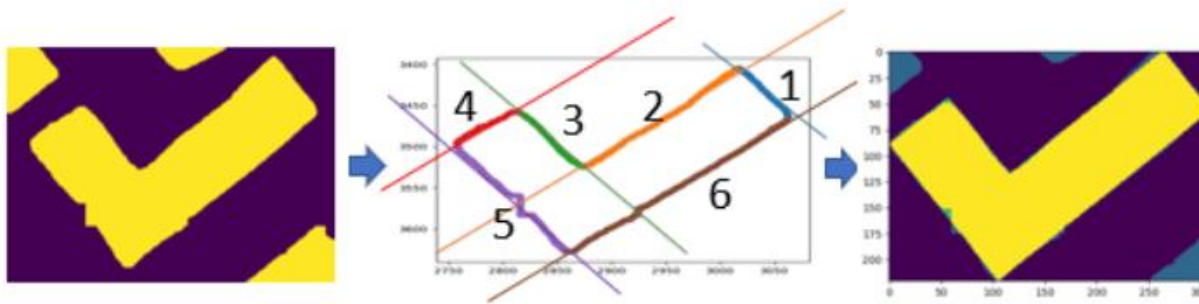
# Microsoft: 125 миллионов полигонов зданий

Автоматическое картирование зданий

## First Stage - Semantic Segmentation



## Second stage - Polygonization



<https://github.com/microsoft/USBuildingFootprints>

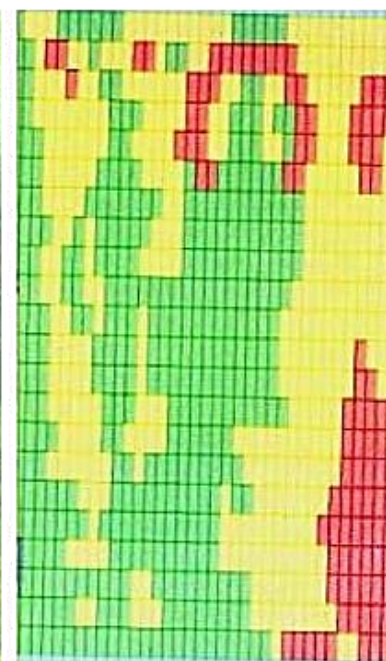
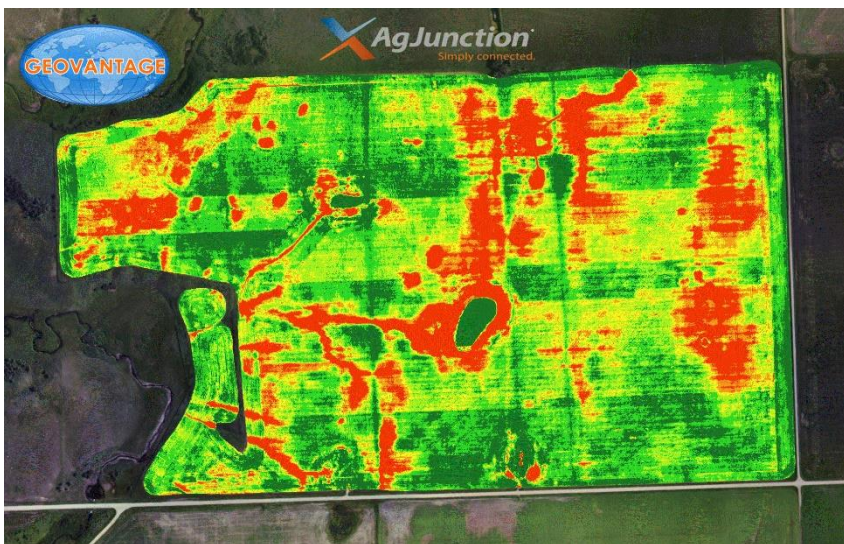
<https://azure.microsoft.com/en-us/blog/how-to-extract-building-footprints-from-satellite-images-using-deep-learning/>

# Точное земледелие/precision agriculture

- Быстрое выявление заболеваний, вредителей
- Точный расчет доз удобрения для каждого участка поля
- Оценка влажности почвы и других характеристик
- Прогноз урожайности для каждого участка поля
- ...

## Современный тренд: данные БПЛА

- стоимость,
- скорость,
- периодичность, ...



Red = 210 kg/ha N  
 Yellow = 180 kg/ha N  
 Green = 130 kg/ha N

# UAV: Unmanned Aerial Vehicles



[http://www.youtube.com/watch?v=4ErEBkj\\_3PY](http://www.youtube.com/watch?v=4ErEBkj_3PY)

# Стартапы

Слайд из презентации TerraTex от 28 апр 2020 г.

## Привлеченные инвестиции в создание новых космических группировок



BLACK SKY  
THE WORLD NOW

**\$175 млн**

Создание к 2020 г. группировки из 60 малых спутников Global (55 кг) для обеспечения оперативной повторной съемки любого участка земной поверхности в течение 90 минут. Спутники будут вести фото- и видеосъемку с разрешением 1 м.

США

Synspective

**\$100 млн**

Создание группировки из 25 небольших (порядка 150 кг) радиолокационных спутников с синтетической апертурой для съемки всей Земли.

Япония

HYPERSAT™

**\$85 млн**

Создание коммерческой группировки гиперспектральных спутников (200+ каналов) весом 200-300 кг с высоким разрешением.

США

HawkEye<sup>360</sup>  
All-Aspect Mobile Company

**\$70 млн**

Создание группировки из 30 пассивных малых радиоспутников (13 кг). Группа малых спутников будет собирать информацию о радиосигналах по всему миру для обеспечения высокоточного радиочастотного картирования и анализа.

США

ICEYE

**\$54 млн**

Создание группировки из 18 радарных микроспутников (менее 100 кг), которые будут оснащены радарными с синтетической апертурой (SAR). Спутники обеспечат оперативную съемку практически в ежечасном режиме.

Финляндия

SATELLOGIC™

Аргентина

**\$50 млн**

Создание группировки из 300 микроспутников NuSat (37 кг). Главное назначение группировки — фото- и видеосъемка Земли для коммерческих целей и общественных нужд практически в реальном режиме времени.



Великобритания

**\$50 млн**

Создание группировки из 15 оптических спутников ДЗЗ (до 100 кг), которые будут вести цветную фото- и видеосъемку высокого разрешения меньше 1 метра в мультиспектральном режиме с возможностью анализировать движущиеся объекты.

NORTHSTAR  
EARTH & SPACE

Канада

**\$40 млн**

Создание группировки из 40 низкоорбитальных спутников (будут летать по орбите, на которой работают спутники связи), предназначенных для наблюдения за окружающей средой.

Capella Space

США

**\$34 млн**

Создание группировки из 30 недорогих радарных (SAR) малых спутников с синтетической апертурой Capella (40 кг). Первая очередь спутников будет иметь разрешение в 1 м.



США

**\$23 млн**

Создание группировки из 30 наноспутников Landmapper (11 кг) для проведения мультиспектральной съемки всей земной поверхности.

# **Research directions & Project Topics**

**the list is not exhaustive!**

# Change Detection

## Обнаружение изменений

### Примеры обнаружения изменений

- на месте здания появилось футбольное поле
- на месте пустыря появился жилой комплекс
- длина строящегося моста выросла на 250 метров за прошедшие пол года
- площадь зеленых зон в городе увеличилась на 10%

### Обзор методов: From pixel-based to object-based approaches [PDF](#)

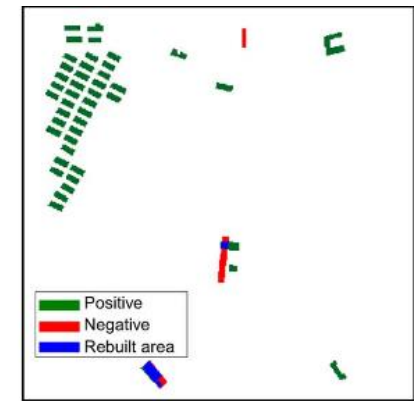
- нейронные сети
- SVM
- деревья решений
- нечеткая логика
- объектные подходы
- ...



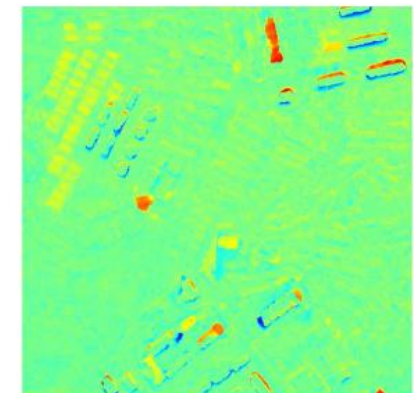
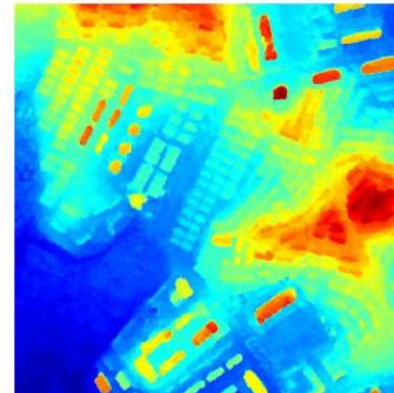
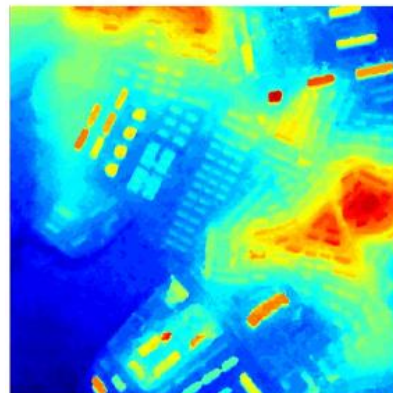
(a)



(b)



(c)

[LINK](#)


# Классификация объектов на спутниковых снимках с помощью глубокого обучения

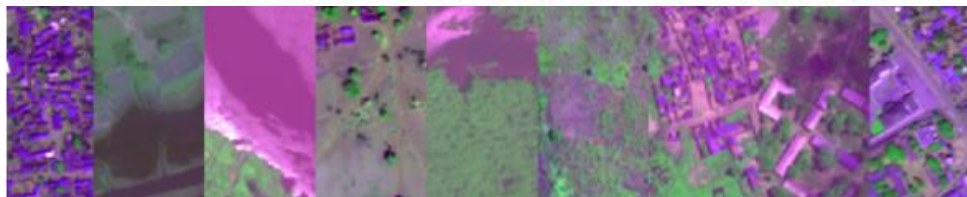
<https://habrahabr.ru/company/avito/blog/325632/>

<https://habrahabr.ru/company/ods/blog/325096/>



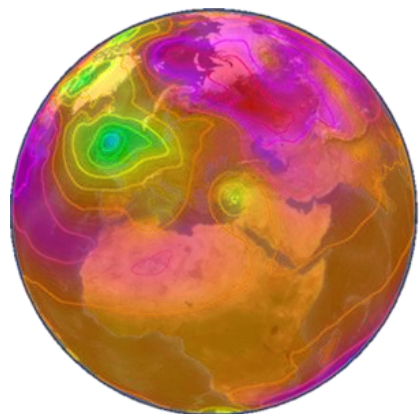
а конкурса Dstl

ритмы, Блог компании Avito

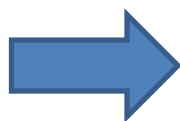


# Содержание работ

- Командный или индивидуальный проект



Изучение  
данных



Импорт  
данных



Изучение системы

- 25 систем на выбор +  
ваши предложения

Программирование  
либо составление  
запросов: с какими  
задачами система  
справляется лучше, а  
с какими хуже



- разное число узлов кластера
- разный объем данных
- модификация запросов для улучшения/ухудшения производительности



# Результаты Вашей работы

## Востребованные на рынке труда компетенции:

- **Опыт работы с гео-данными**

Характер данных, форматы, алгоритмы, особенности обработки гео-данных



- **Опыт работы с системами**

Например, многие системы основаны на Apache Spark, одной из самых востребованных систем работы с Большими Данными



- **Опыт исследовательской работы**

- **Публикации (по желанию)**

Преимущество в трудоустройстве (full-time/intern), конкурсе НИРС, повышенной стипендии, поступлении в магистратуру



# Microsoft, IBM, Intel, Nvidia, Yandex, ...

[Microsoft 365](#)[Azure](#)[Office 365](#)[Dynamics 365](#)[SQL](#)[Windows 10](#)[More](#) ▾[Research](#)[Research areas](#) ▾[Products & Downloads](#)[Programs & Events](#) ▾[People](#)[Careers](#)[Blogs & Podcasts](#) ▾[Labs & Lo](#) >

## Fuel the future

Join a brilliant team of researchers working to solve technology's most exciting challenges.



### Full time

<https://www.microsoft.com/en-us/research/careers/>

Microsoft works with the world's best **researchers**, who move with the current of technology as it rapidly evolves. Together we share the motivation to seek innovative solutions to the world's toughest challenges and improve the lives of people everywhere. We seek research candidates with PhDs and **a proven track record of published papers** and participation on program committees,

#### Success story: Евгений Поздеев

- **2 статьи на конференциях**
- **двукратный победитель конкурса НИРС**
- **Июнь 2017 – Сентябрь 2017: стажировка в Google, Дублин, Ирландия**
- **Сейчас: магистрант в Университете Брюсселя**

Program for Target Recognition on Satellite Images

Program for Satellite Scene Understanding

Research and Development of Methods for Approximate Query Answering in Array Databases using Neural Networks

Software Accelerator for the Virtual NcML Format using GraalVM

Research and Development of Methods for Multidimensional Array Processing on GPU

Research and Development of Methods for Vector Geospatial Data Processing on GPU

Research and Development of Methods for Satellite Data Processing on GPU

Performance Evaluation of Apache Spark on Raster Geospatial Data

Performance Evaluation of Apache Spark on Vector Geospatial Data

Performance Evaluation of MrGeo on Raster Geospatial Data

Performance Evaluation of Julia Programming Language on Geospatial Data

Version Control System for Geospatial Data Supporting Analytical Queries

Research and Development of Elastic Big Raster Data Processing Algorithms

Research and Development of Methods for Urban Change Detection from Satellite Data Based on Neural Networks

Research and Development of Methods for Land-Use Change Detection using Multisensor Satellite Data

Research and Development of Methods for Change Detection in Synthetic Aperture Radar Images Based on Deep Learning

Research and Development of Methods for Patterns Discovery in Geospatial Data using Data Mining

Research and Development of Methods for Deep Learning-Based Classification of Hyperspectral Data

Research and Development of Methods for Analysis of Big Spatio-Temporal Data Volumes Using the Approximate Structures

Research and Development of Methods for Satellite Imagery Feature Detection Using Machine Learning

Research and Development of Methods for Patterns Discovery of Objects Tracks using Data Mining

Distributed Software System for Land-Use Change Detection using Multisensor Satellite Data

Performance Evaluation of GeoMesa on Geospatial Data

Performance Evaluation of GeoTrellis on Raster Geospatial Data

Performance Evaluation of GeoTrellis on Vector Geospatial Data

Performance Evaluation of GIS Tools for Hadoop on Geospatial Data

Performance Evaluation of Spatial Hadoop on Geospatial Data

Performance Evaluation of Hadoop-GIS on Geospatial Data

Performance Evaluation of GeoSpark on Geospatial Data

Performance Evaluation of Magellan on Geospatial Data

Performance Evaluation of Julia Programming Language on Geospatial Data

Performance Evaluation of PostgreSQL/PostGIS on Geospatial Data

Performance Evaluation of SciDB on Geospatial Data

Performance Evaluation of RasDaMan on Geospatial Data

Research and Development of Methods for Indexing Big Satellite Data

Research and Development of Big Raster Data Caching Methods

Research and Development of Methods to Perform Map Algebra Operations using GraalVM

# Topics in LMS

**if you are interested in R&D in the  
area of geoinformatics**

mail to [arodrigues@hse.ru](mailto:arodrigues@hse.ru)

**and we will discuss**

**1st year** (course work) and **2nd year** (M.S. Thesis)

**projects that you can work on**

**(even if they are not yet in LMS)**



NATIONAL RESEARCH  
UNIVERSITY

Thank you  
for your attention!