

INFOMDW: Data Wrangling

Introduction to the course

Hakim Qahtan

Department of Information and Computing Sciences

Utrecht University



Utrecht University

Today

- Who are we?
 - Introduction to the course content
 - Getting to know you a little
 - Break
 - Practicalities & course flow
 - Time for questions
-
- But first...



Who are we?



The Teaching team

Hakim Qahtan
a.qahtan@uu.nl



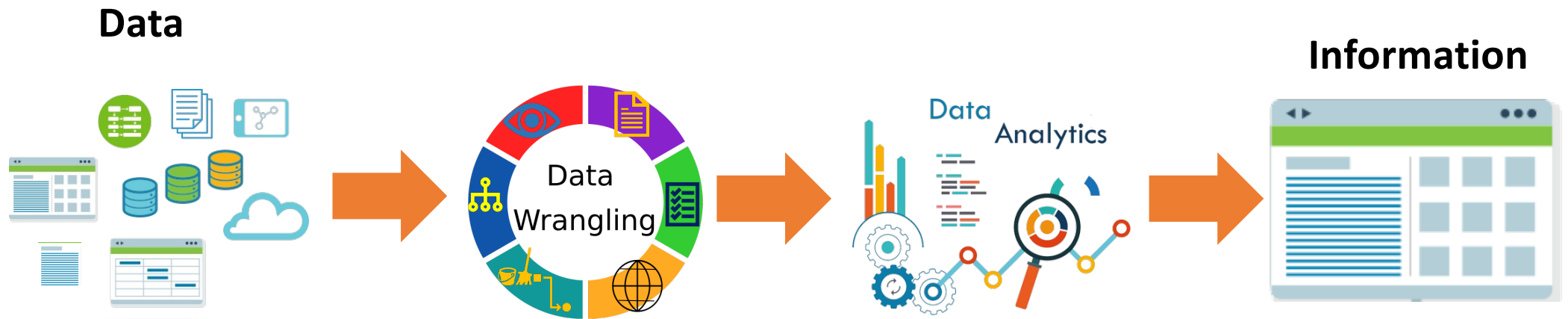
Konstantinos Nikoletos
k.nikoletos@uu.nl



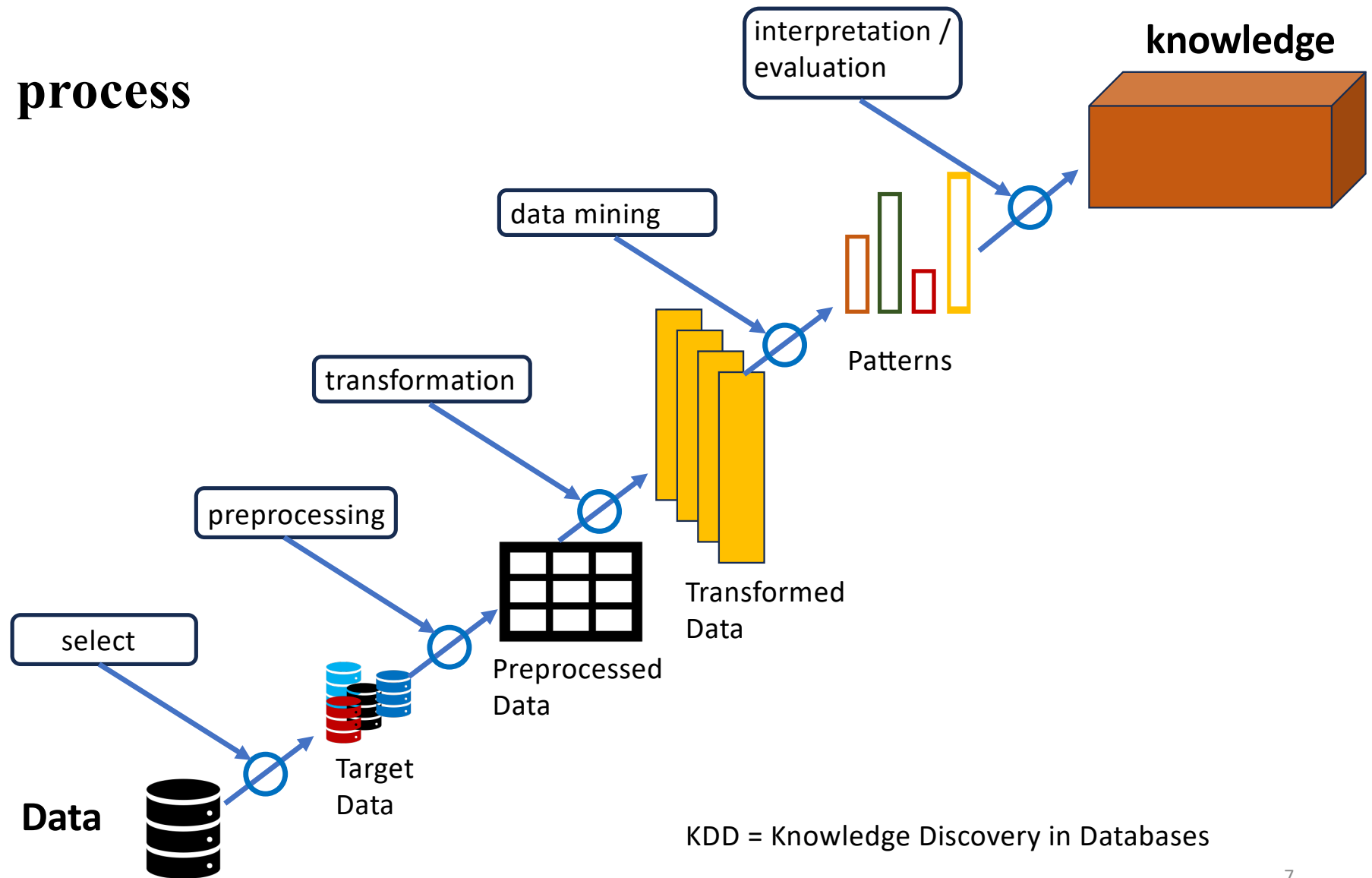
Course introduction



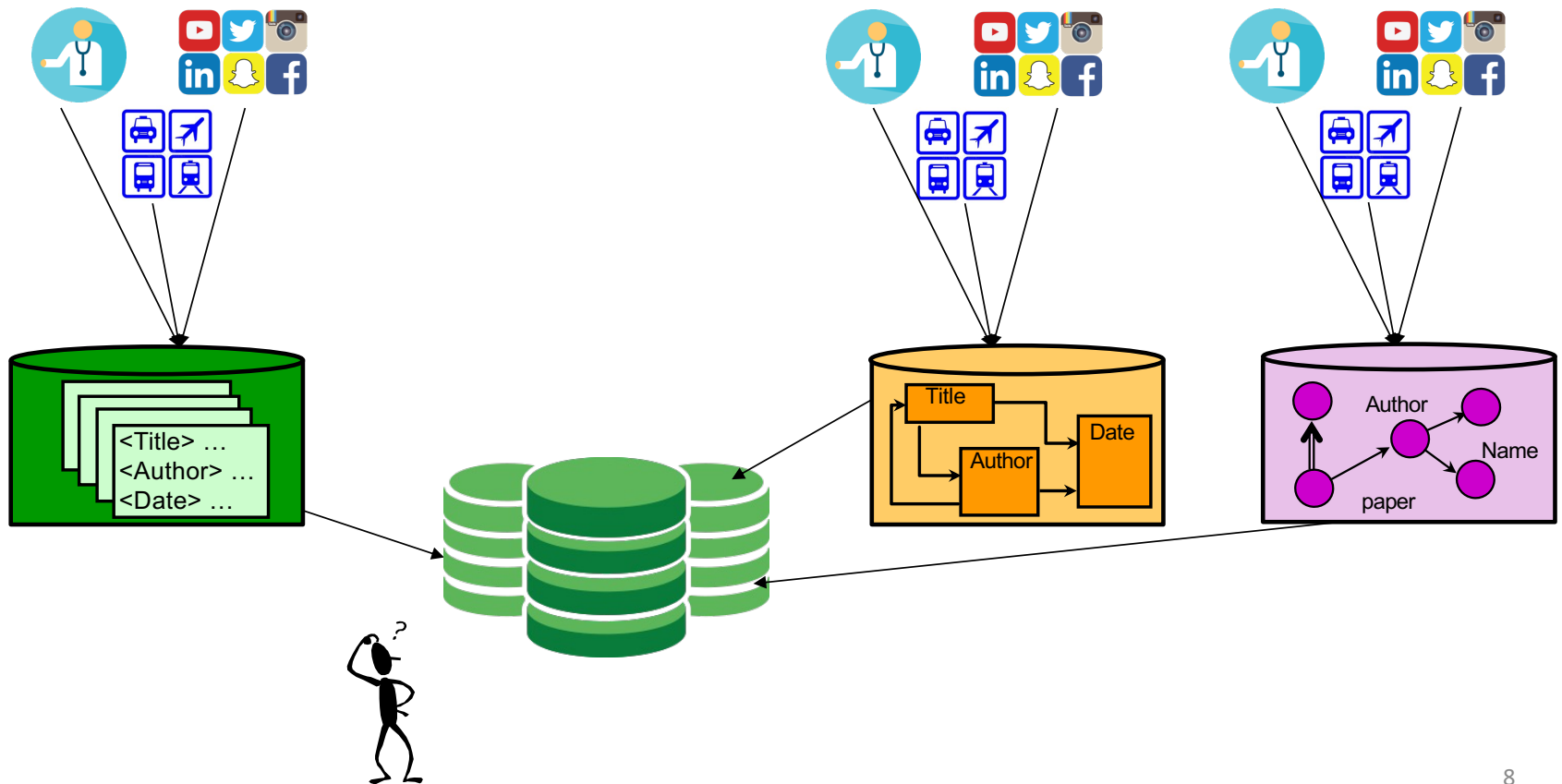
Extracting value from data



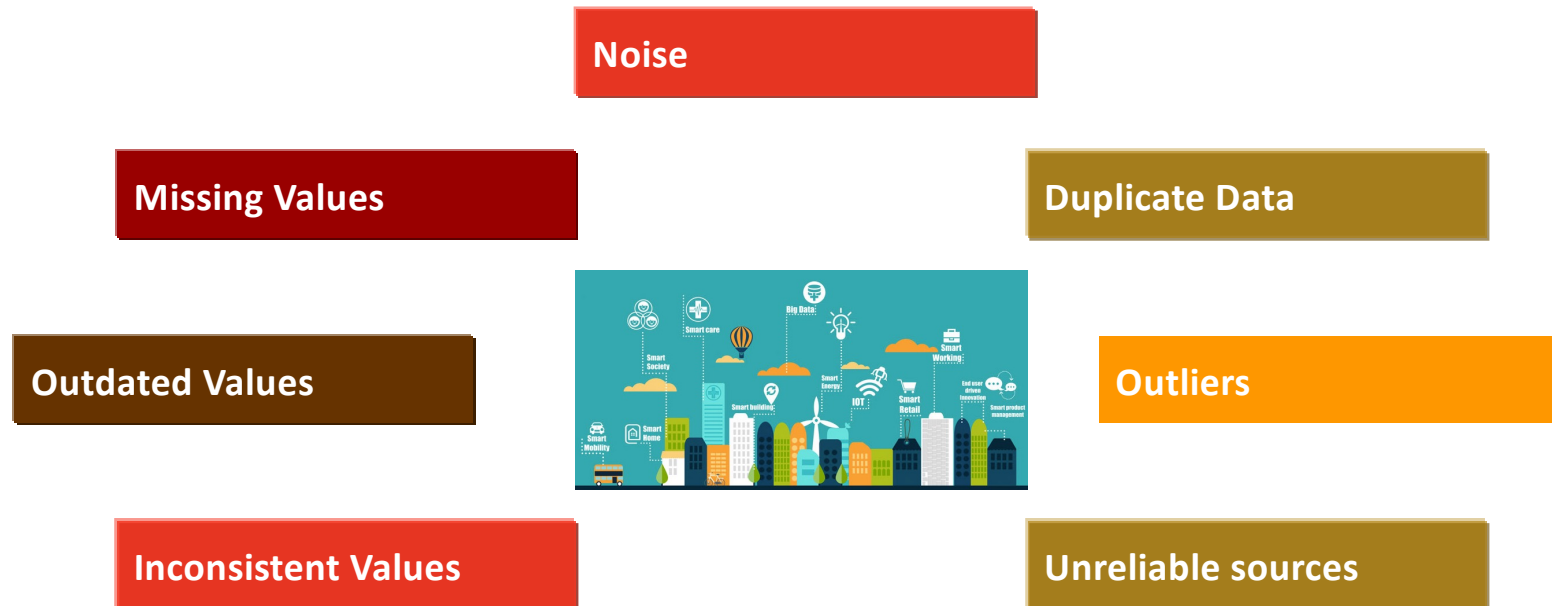
KDD process



We collect a lot of data



Data quality issues

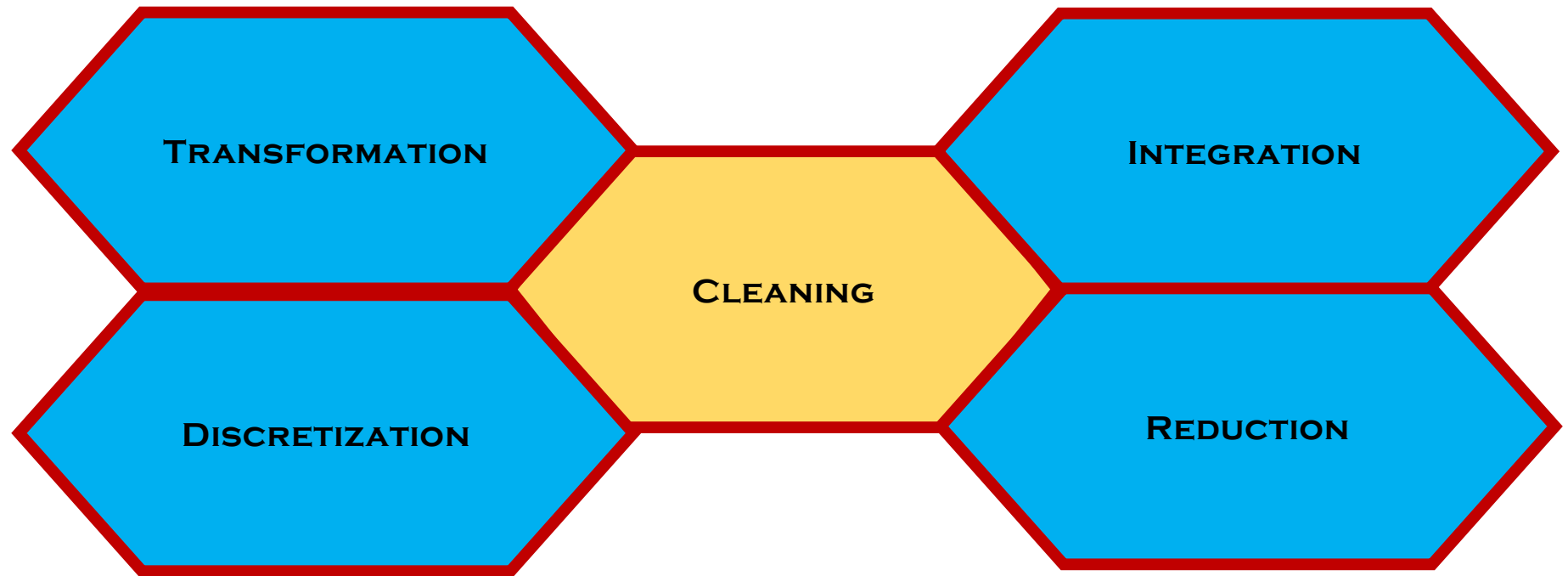


For any data retrieval or data analytics task,
it is critically important to know the quality of your data

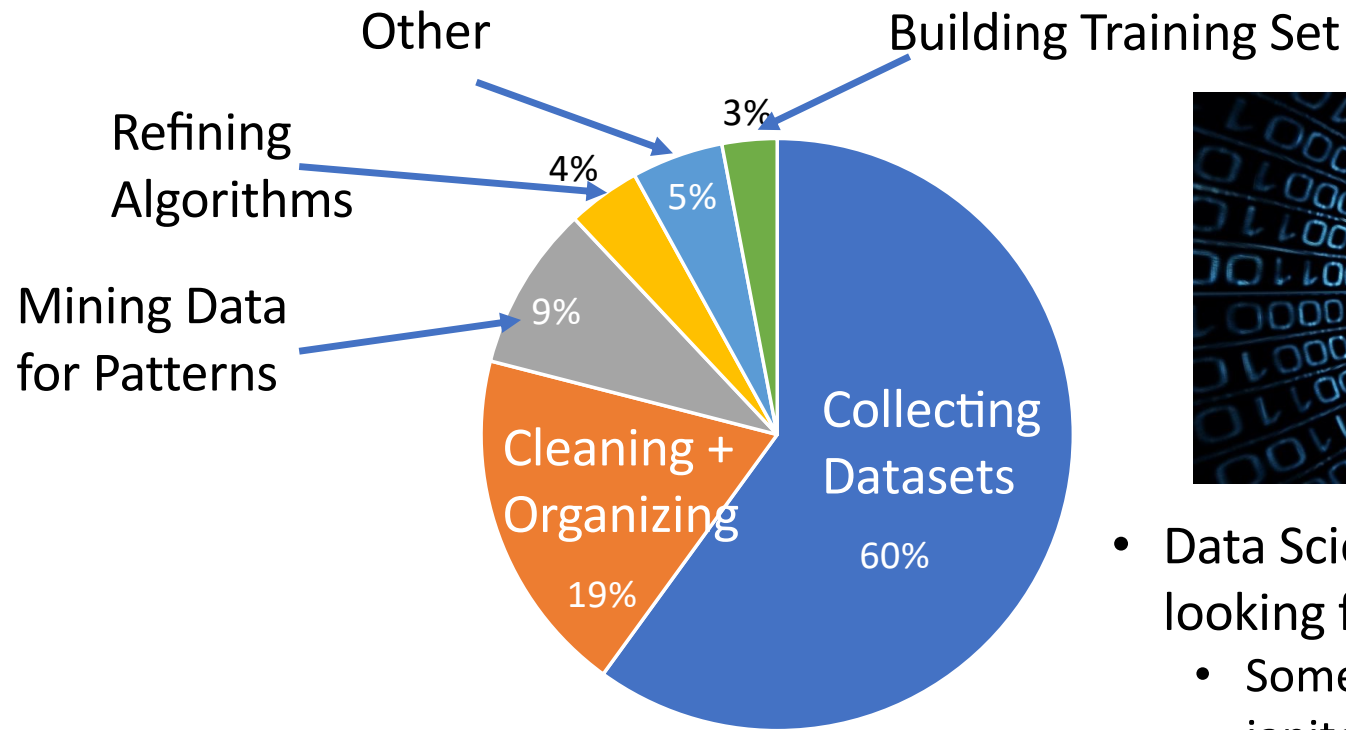
For the US alone, poor data quality costs around US economy \$3 trillions a year (2016).



Data preparation



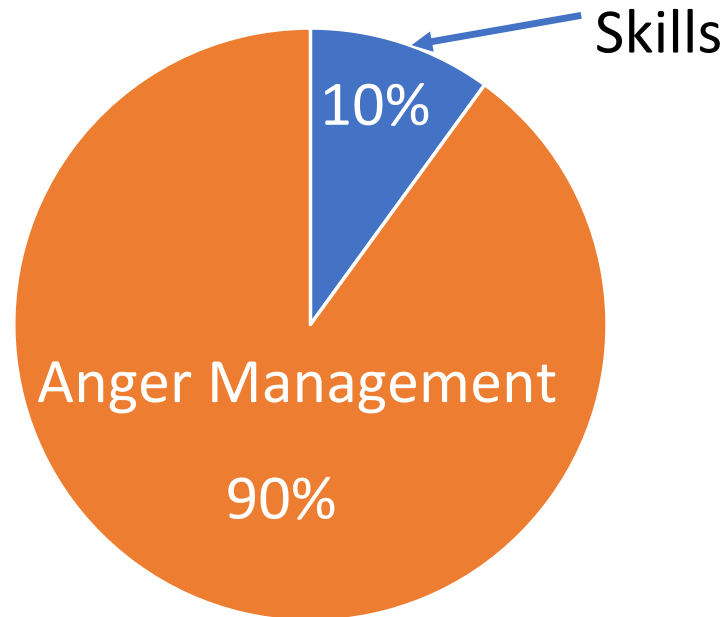
How are you going to spend your time?



- Data Scientists: 80% of their time looking for and cleaning the data
 - Sometimes, they are called data janitors

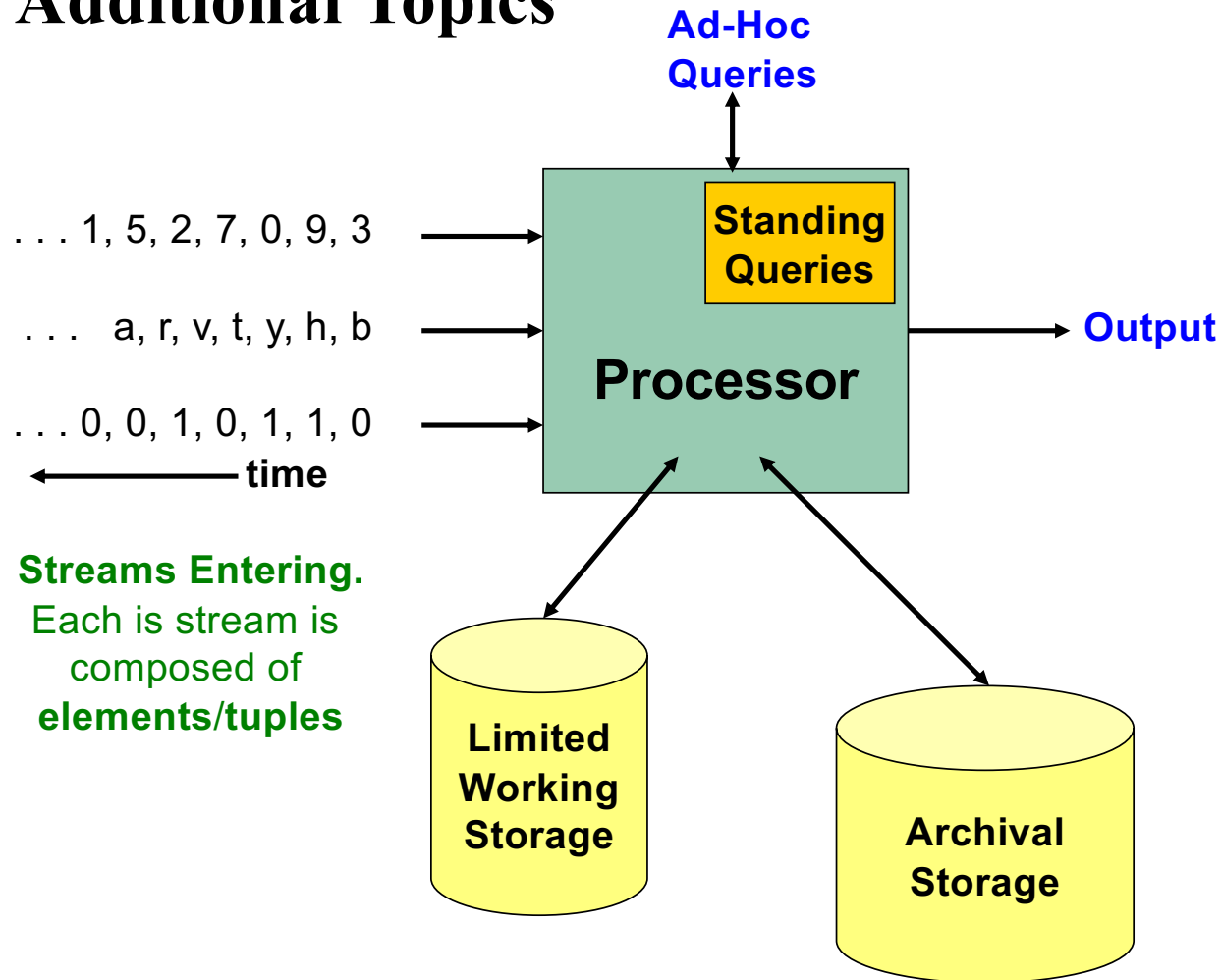


Data Wrangling

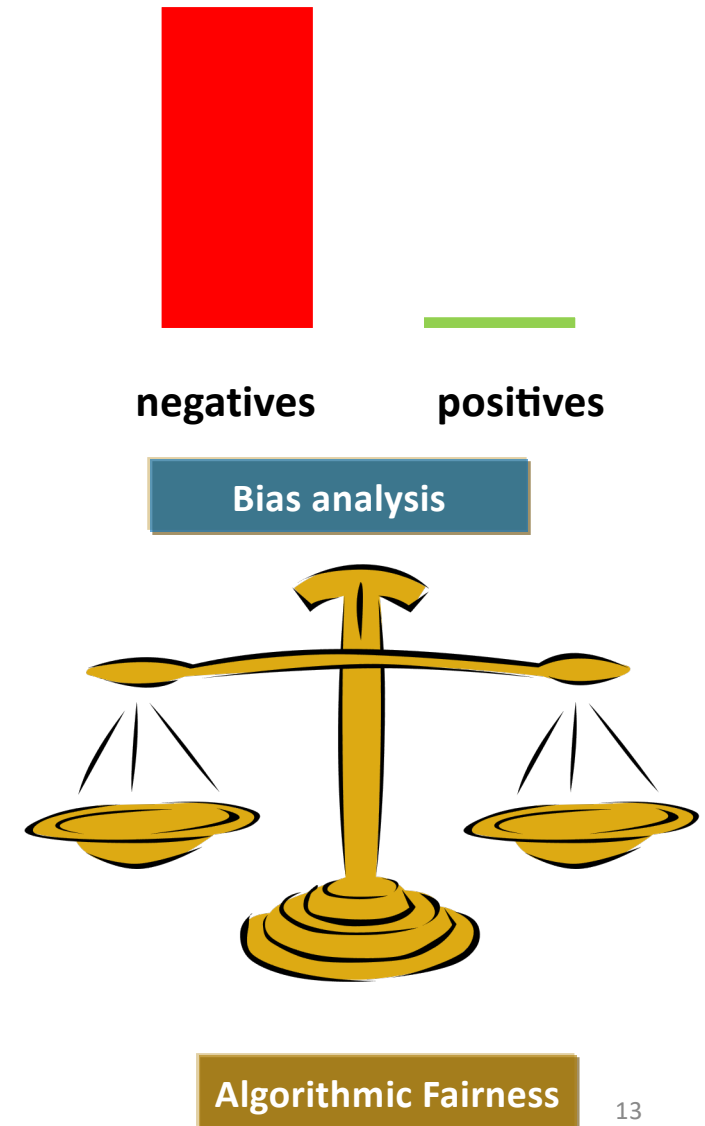


PHIADELPHIA
PHIALDELPHIA
PHIDELPHIA
PHIELADELPHIA
PHIILADELPHIA
PHILA
PHILA.
PHILAD
PHILADALPHIA
PHILADEDLPHIA
PHILADELAPHIA
PHILADELHIA
PHILADELHPIA
PHILADELLPHIA
PHILADELOHIA
PHILADELPH
PHILADELPHA
PHILADELPHAI
PHILADELPHI
PHILADELPHIA
PHILADELPHIA PA
PHILADELPHIA,
PHILADELPHIA, PA
PHILADELPHIA'
PHILADELPHIAP
PHILADELPHIAPHIA
PHILADELPHILA
PHILADELPHIOA
PHILADELPIA
PHILADELPOHIA
PHILADELPPHIA
PHILADEPHA
PHILADEPHIA
PHILADEPHILA
PHILADEPLHIA
PHILADERLPHIA
PHILADELPHIA
PHILADLEPHIA
PHILADLPHIA
PHILADPHIA
PHILADRLPHIA
PHILAEPLHIA
PHILDADLPHIA
PHILDAELPHIA
PHILDELPHIA
PHILDEPPHIA
PHILIADELPHIA
PHILIDELPHIA
PHILLA
PHILLADELPHIA
PHILLY
PHILOADELPHIA
PHLADELPHIA
PHOLADELPHIA 12
PHPIADELPHIA
PIHLADELPHIA

Additional Topics

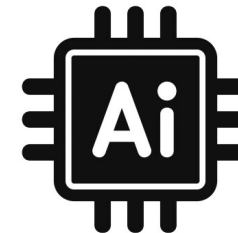
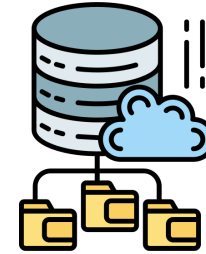


Streams Entering.
Each is stream is
composed of
elements/tuples



This course is not

- Database course
- Systems course
- Artificial intelligence course
- Machine learning course





Go to
wooclap.com

Enter the
event code
in the top
banner

Event code

DW2627



<https://app.wooclap.com/DW2627>



Break



Practicalities

<https://hqahtan.github.io/infomdw.html>



Practicalities

- Everything is on the course website <https://hqahtan.github.io/infomdw.html>, except:
 - Room locations (on mytimetable.uu.nl)
 - Announcements (in your email and on uu.brightspace.com)
 - Assignment hand-in (on uu.brightspace.com)
 - Exams (on the university digital assessment tool Remindo)
 - Lectures / tutorials will be in person



materials on the website are constantly under construction



Course flow

- This is a mandatory course.
- Every* week you will:



Read the required readings



Attend the lectures: Tuesday, Thursday



Attend and work on the hands-on tutorials: Thursday



Work on the group and the individual assignments



Review materials diligently for the midterm and final exams

* some weeks contain other stuff like exams and study time



Schedule & readings

- The course website contains a full schedule with required readings
- Reading materials can be found online for free! 🕵️ (you are resourceful students)



Lectures: Prepare for next lecture!

1. Look up reading materials in course website
2. Look up time and location for the lecture on mytimetable or the university app



Tutorials

- Hands-on tutorials with practical exercises
- Put into practice what you learnt in the lectures
- We incorporate real-world data and use-cases: APPLIED data science!
- One session per week with practical exercises from the theoretical lectures
- **The teaching assistant is your main point of contact for questions!**
- Tutorials provide skills needed to do the assignments



Assignments

- Four assignments in total
 - First assignment is already available on the course website
 - deadlines: weeks 4 and 8 for Group assignments and weeks 7 and 10 for individual assignments
- Hand in on Brightspace
- For group assignments!
 - You can assign yourself in a group starting from Friday (tomorrow) until Wednesday next week.
- **Plan ahead**



Assignments: prepare!

1. When is the deadline of the first assignment?
 1. Check the course website
2. Quickly check the content of the first assignment
3. Log in to Brightspace and find the INFOMDW course
4. If everything is well organized:
 1. Find the group sign-up page there to see how you can sign up
 2. Find the assignment hand-in page to see how you can submit
5. If there are any problems, don't worry, send us an email and it'll be fixed soon 😊



Announcements

ADS OAC is looking for Student Members

- ADS Education Advisory Committee (OAC) is looking for three student members
 - One member (vice-chair) will be a member of GSNS OC
- OAC is composed of equal number of staff and student members.



What will you do?

- You will:

- Review the Caracal evaluations
- Comment and give suggestions on the GSNS OER (Education and Examination Rules) and program-wise (ADS) OER Annex
- Help in solving other education quality-related issues

- Expected time dedication

- Weekly average: 2 hours for OAC members, 4 hours for vice-chair (also member of GSNS-OC)
- You will be compensated for the time you spent



How to apply?

- Interested?



+



Angelos Chatzimparmpas
a.chatzimparmpas@uu.nl



ADS Workshop

- Date: 5 November
- Location: KBG-COSMOS and KBG-PANGAEA
- Time: 9:00 – 16:45
- 4 guest speakers + 1 Q&A session
- Add the event to your calendar



Further questions?



See you in the tutorial session 🖐️
First session is optional!

