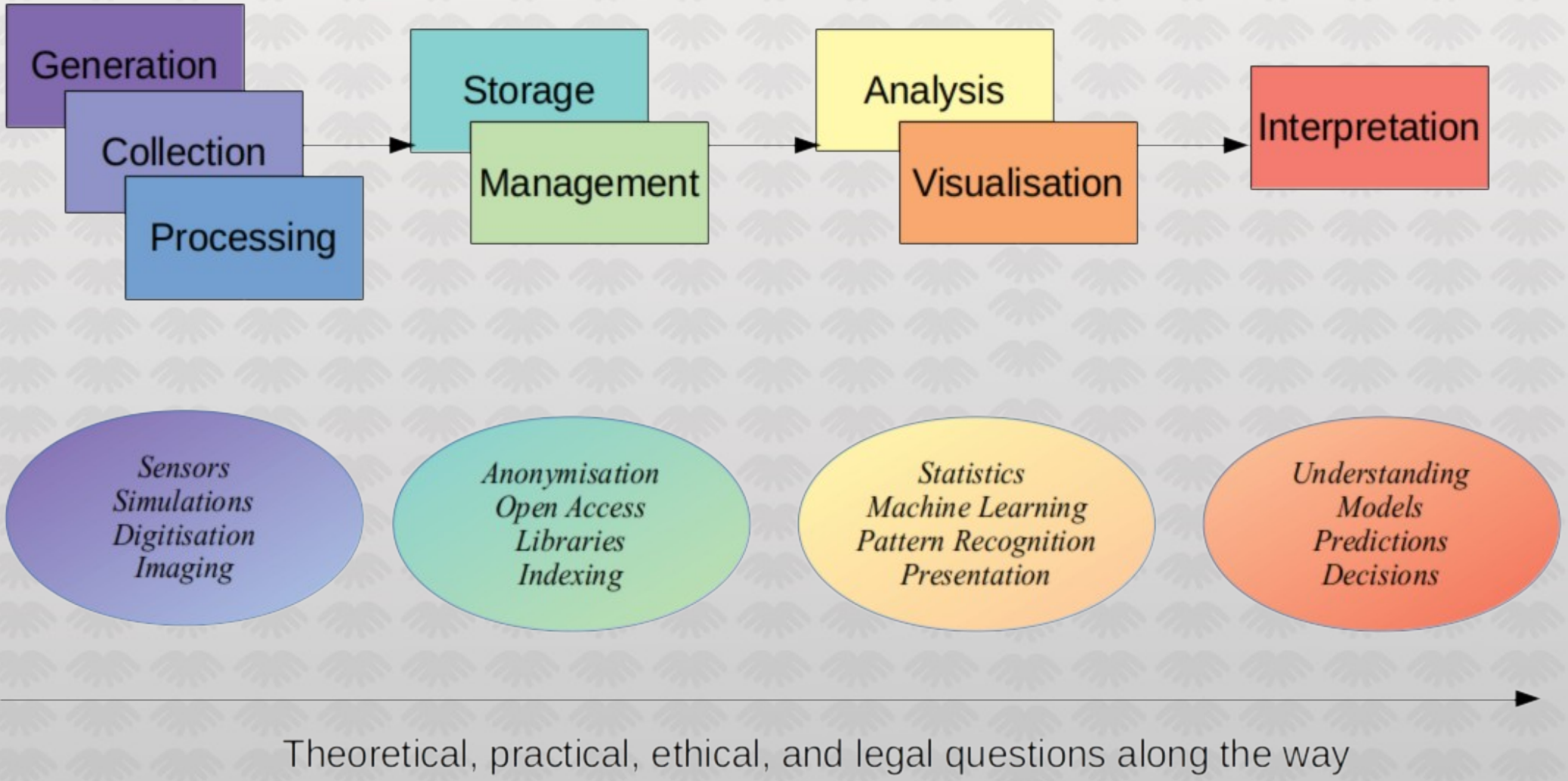
Durham University

# IDAS – Institute for Data Science

## mission & plans

**Frank Krauss**
**Institute for Data Science**
**Durham University**

# Mission and plans


Generation
Collection
Processing
Storage
Management
Analysis
Visualisation
Interpretation
Sensors
Simulations
Digitisation
Imaging
Anonymisation
Open Access
Libraries
Indexing
Statistics
Machine Learning
Pattern Recognition
Presentation
Understanding
Models
Predictions
Decisions
Theoretical, practical, ethical, and legal questions along the way

**IDAS' mission** is the provision of an umbrella for Data Science at Durham University, by

- fostering multi-disciplinary research:
  - developing new research themes and formulating novel ideas
  - connecting essential methodological and technical skills with critical domain knowledge,
  - promoting best practice
- answering to societal and economic trends
  - responding to Data Science questions by local, regional and national private and public sector actors and agencies
  - providing a platform for applications for external funding from multiple sources

| Language | Images | Numerics | Culture |
| --- | --- | --- | --- |
| Natural Language Processing | Imaging of artefacts and the physical environment | Statistics, Algorithms and HPC | Ethics, Law and Philosophy |

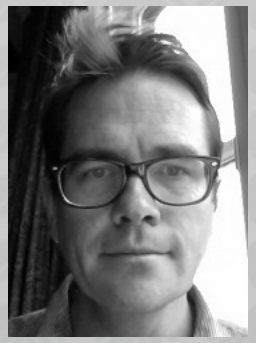
Digital Humanities

Scientific and Environmental Data

Economics of the Future

Development, Behaviour and Interactions

**Technological Backbone: ARC**



# IDAS invites bids for interdisciplinary projects

- application with less than 2,500 words
  - who: track record of applicants
  - what: project context, outcomes, deliverables
  - why: strategic beenfit for IDAS & University
  - how: up to £5,000 per project for undergraduate work, material, etc.
- deadline for applications: 16.10., IDAS management board takes decision by 15.11.

# IDAS activities:

- monthly “wine & cheese” colloquia
  - mainly external speakers
  - to take place on 4-5 Friday afternoon (probably here)
- drop-in data sessions
  - short presentation by internal speakers
  - forum to discuss ideas, ask for help
  - to be organised by students
- support for 1-2 workshops per year
  - please contact the subject leads
- advertise your research through the web-page
  - send us the doi / arXiV / … identifier

# Research-Led Training and Education

# Doctoral Training

- **STFC-funded Centre for Doctoral Training in Data-Intensive Science**

  **22 PhD students in Cosmology, Observational Astronomy, and Particle Physics Phenomenology**

  **15 students funded by STFC, 5 by University, 2 self-funded**

  **(2 cohorts, intakes 2017 & 2018)**

  **projects on producing and/or analysing large datasets, mandatory 6 months placement outside academia**

  **(so far: Boeing, P&G, United Nations)**

  **structured first year training, including 8-weeks team project with private or public sector partner or in academia, leading to further opportunities (e.g. ERDF PhD studentship with IBEX)**

  **(last year: 2 in-house, INSPIRE, P&G, IBEX**
  **this year: IBEX, Tharsus, Nissan, Northumbrian Water)**

# CDT Project

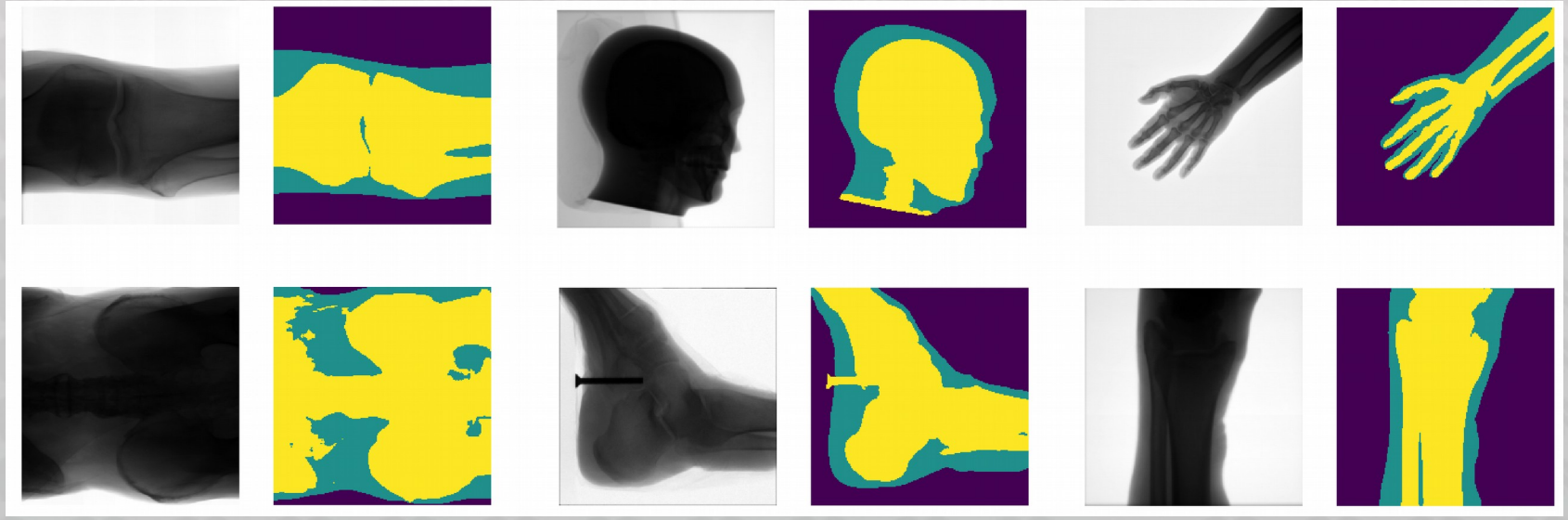
- **8-weeks team project with IBEX (X-ray imaging)**

(Carolina Cuesta-Lazaro, Arnan Quera-Bofarul, Joseph Bullock)

**train NN to distinguish soft tissue and bone**

**students taught Python and NN, results of their project now part of IBEX's software package**

**publication, invitation to largest X-ray imaging conference, winner of poster competition**



# New MSc Programme

- **Scientific Computing and Data Analysis**
  - built on experience from structured 1st year training in CDT, participating departments: CS, Maths, Physics
  - start 2019, about 40 students as first intake
  - modular structure to extend programme:
    - add another “core” stream for social sciences/art & humanities
    - add more specialisations (Health & Finance under discussion)

| | Term 1 (Michaelmas Term) | Term 2 (Epiphany Term) | Term 3 (Easter Term) |
|---|---|---|---|
| **Professional Skills** | • Introduction to Python<br>• Software Carpentry<br>• Introduction to Software Project Management<br>• Introduction to Scientific Writing | • Data Presentation and Visualisation<br>• Versioning Systems<br>• Software Project Management<br>• Writing of Technical & Scientific Reports | • Communicating Science<br>• Mini-MBA |
| **Scientific Computing & Data Analysis** | **Core I: Statistics, Machine Learning, Scientific and High-Performance Computing**<br>• Introduction to Data Analysis<br>• Introduction to Scientific Computing<br>• Introduction to High-Performance Computing<br>• Introduction to Machine Learning | **Core II A: Advanced Statistics and Machine Learning**<br>• Advanced Data Analysis<br>• Advanced Machine Learning<br>• Introduction to Data Acquisition<br>**Core IIB: Advanced Scientific and High-Performance Computing**<br>• Advanced High-Performance Computing<br>• Advanced Scientific Computing | **Project**<br>• Duration: 8-10 weeks<br>• Written thesis<br>• Options:<br>◦ With public or private sector partner, typically as part of a team of 2-3 students;<br>◦ Within subject specialisation; or<br>◦ Methodological work |
| **Subject Specialisation: Astronomy** | | | |
| **Subject Specialisation: Particle Physics** | • Introduction to Field Theory<br>• Introduction to Quantum Field Theory<br>• Quantum Electrodynamics<br>• Standard Model | • Quantum Chromodynamics<br>• Effective Field Theories & Flavour Physics<br>• Neutrinos & Astroparticle Physics<br>• Higgs Physics | |