CODATA/RDA Summer schools in Data Science and Cloud computing for research in the developing world

KIT
12 June 2015

Vision

Create a sustainable series of Summer Schools in Data Science and Cloud Computing for early career researchers. Support for researchers from the Developing World.

Provide pathways as there are a variety of different roles/careers in Data Science

Medium to longer term goal - provide rigorous accreditation

That's nice - but why ?

Plenty of online material / MOOCs

Multiple schools/bootcamps

Accreditation also exists

Reasons

Resources for Researchers in Data Science are scattered

Courses are oriented towards business

Longer term goal - win the Open Data argument with the young

Peace, Love and Data

Enormous enthusiasm within RDA

Data centres, funders, senior researchers “get it”

Genuine interest in enabling open data

Understanding that open data enables

Research Reproducibility Careers

On the other hand….

Much less interest from research community

Within UK - Data storage requirements from

BBSRC
EPSRC
NERC

Purely Anecdotal - Seen as impositions rather than enabling

It is hard work persuading haggard cynical 40-something researchers like this.

Driving force -
Where's my next grant?
Where's the next paper?
How does this push my career forward?

As Max Planck put it very bluntly…

"A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it."

OR…

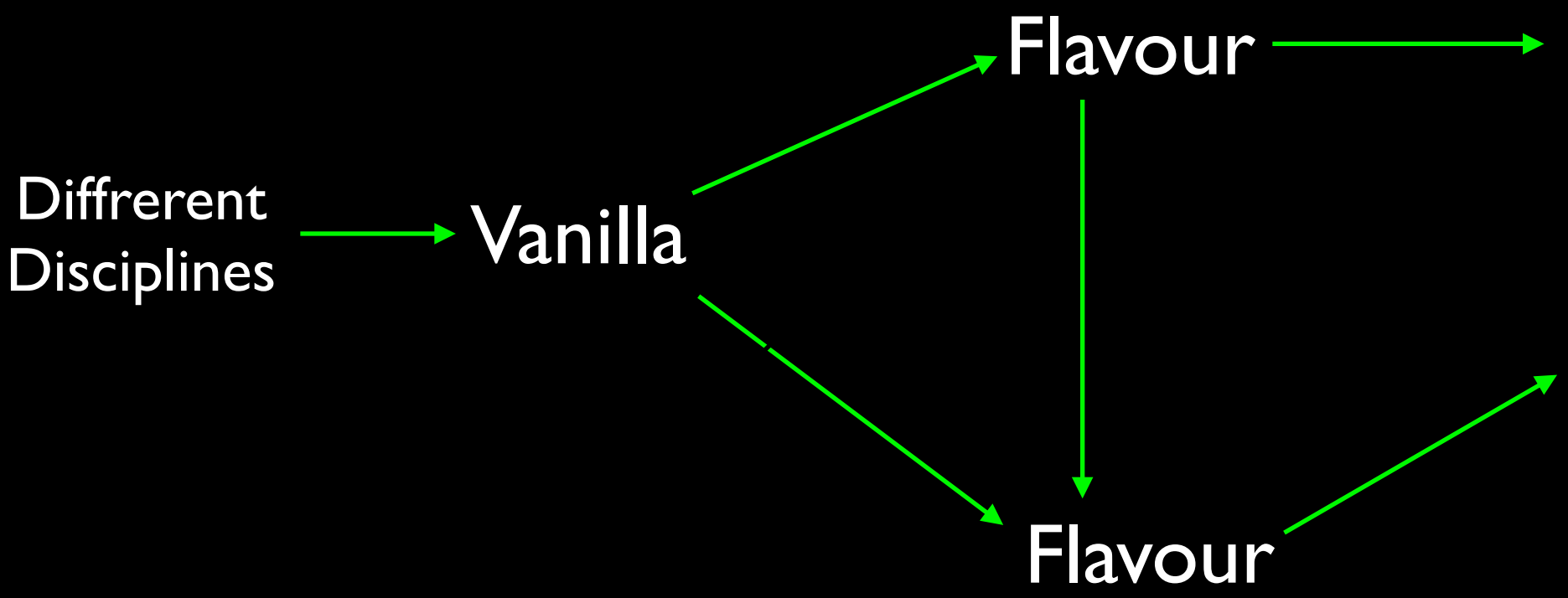
"I believe the ~~children~~ early career researchers are our future."

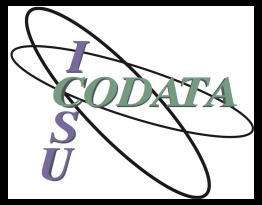
Pace Whitney Houston

Structure

- Intake students at recent graduate student level.
- “Vanilla” schools provides a proven ground level understanding.
- “Flavoured” schools cover more advanced topics.
- Each school runs for two weeks
- Each school will have a train-the-trainers session
- All teaching materials will have a CC-BY licence

Progression





Vanilla curriculum

Statistics/Machine Learning

Software Carpentry

Data Carpentry

Infrastructures

Visualisation

Implementation

First Vanilla school will definitely run August 1-12 2016
Trieste, Italy, ICTP

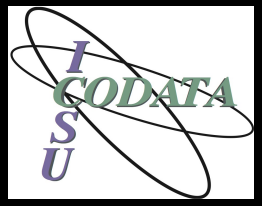
Initial funding from ICTP, TWAS and CODATA

Up to 200 attendees

More Vanilla events

Dry run likely in November 2015 (Cape Town)

and possibly March/April 2016 (Morocco)



Flavours

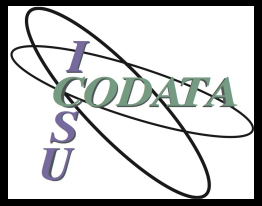
- Covering more advanced topics
- Assumes that all have done Vanilla
- More oriented towards domains, but not entirely.

Extreme Data

Based on analysing very large data sets being generated at facilities such as CERN and the SKA

SKA will generate 2 Tbyte of data *per second*

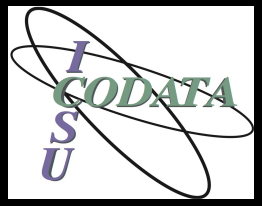
Extreme Data Partners

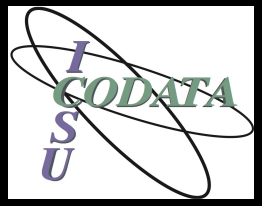


Biological Data

GOBLET
Global Organisation for Bioinformatics
Learning, Education & Training

Geospatial Data





Database Exotica

JACOBS
UNIVERSITY

Library Science

Ligue des Bibliothèques Européennes de Recherche
Association of European Research Libraries

LIBER is Europe's largest network of research libraries, with over 400 members.

Confederation of Open Access Repositories

PURDUE UNIVERSITY LIBRARIES

SUB | NIEDERSÄCHSISCHE STAATS- UND UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN
Georg-August-Universität Göttingen

Agricultural Science

Food and Agriculture Organization of the United Nations

Developing World Science

Cape Town contender for holding initial flavoured schools

Comes from IG on Developing World Research

Commitment to Data Science from South Africa on the basis of the SKA

Could be rolled out to run anywhere

Data Festival

Run flavoured schools sequentially in same location

Students can attend more than one school

Sharing of ideas between lecturers

Festivals can move..

Medium/Long term

Can always give a certificate for attendance

Vanilla and two flavours gives enough teaching for a Masters level qualification

A project would round off things for a Masters

Project

Students would be doing a PhD/Postdoc in their topic.

Project would be based on domain.

Minimise destruction of trees

Have to show that it's contributed to the community

Interactions with RDA

Liasing with Education and Training IG

Liasing with members of Metadata grouping (DCC)

Consulting with members of RDA TAB
on content of Vanilla TAB

Consulting with Big Data,
Geospatial, Agricultural IG's

Moving Forward with the RDA and CODATA

Further advice and consultation on content

Additional Flavoured Schools

Authority that comes from a community