UNIVERSITY OF BIRMINGHAM | RESEARCH SOFTWARE GROUP

Software, Theology, High Performance Computing, AI and RSEConUK 2019

Andrew Edmondson
University of Birmingham

What I do… and where I do it…

UNIVERSITY OF BIRMINGHAM | RESEARCH SOFTWARE GROUP

Research Software Group Leader

UNIVERSITY OF BIRMINGHAM | RESEARCH SOFTWARE GROUP

BEAR Software is a range of services provided to researchers, and research groups, with the aim of improving the research software written and used by the researchers. BEAR Software's mission is summed up in the words of the Software Sustainability Institute: '*better software, better research*'.

Part-time PhD

"An Analysis of the Coherence-Based Genealogical Method using Phylogenetics"

UNIVERSITY OF BIRMINGHAM | RESEARCH SOFTWARE GROUP

RSEConUK 2019 Programme Chair

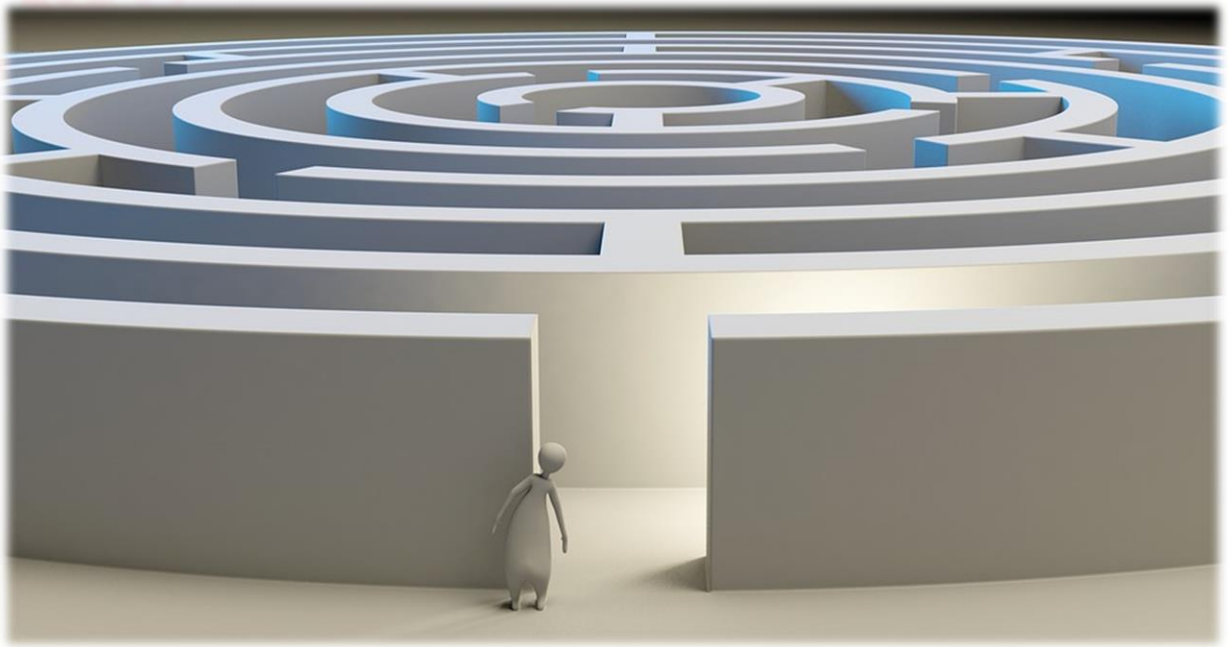
@RSEConUK

UNIVERSITY OF BIRMINGHAM | RESEARCH SOFTWARE GROUP

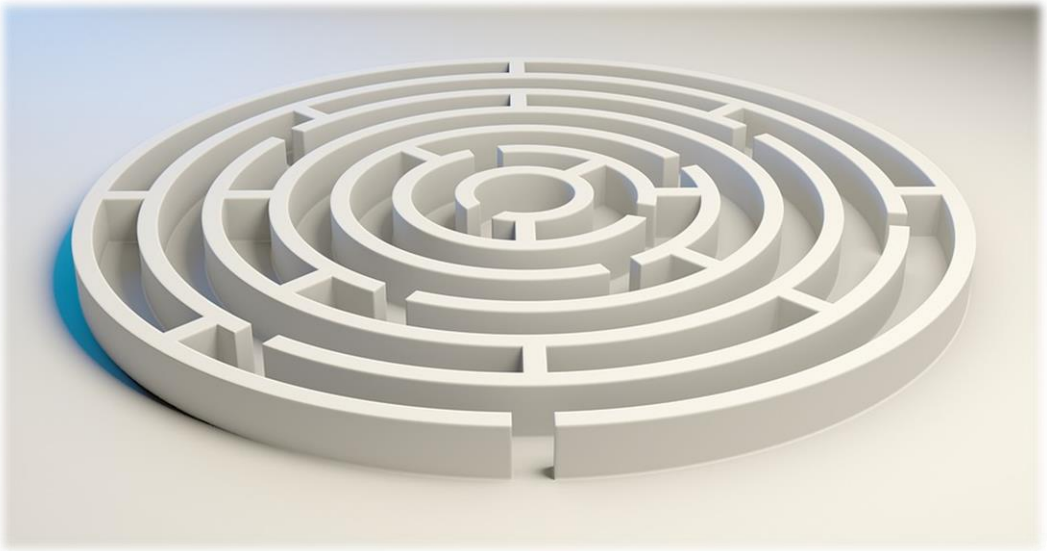
Links

- **BEAR:**
  - https://www.birmingham.ac.uk/bear
  - https://www.birmingham.ac.uk/bear-software
- **RSEConUK 2019:**
  - @RSEConUK
  - https://rse.ac.uk/conf2019
- **Me:**
  - @ed_mondson
  - https://bham.academia.edu/AndrewEdmondson

Introduction



UNIVERSITY OF BIRMINGHAM | RESEARCH SOFTWARE GROUP



University of Birmingham

RSEConUK 2019

Manuscripts and Phylogenetics

University of Birmingham

- Royal charter in 1900 (history back to 1828)
- Member of Russell Group
- 34835 students (2017/18) – 4th largest in UK
- 11 staff and alumni are Nobel prize winners
- £134.2M research income (2017/18)
- £3.5bn economic impact
- Campuses in Birmingham and Dubai

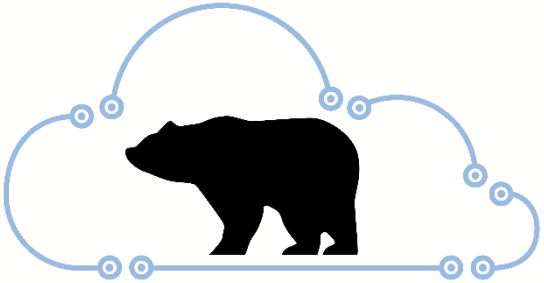
BEAR

BIRMINGHAM ENVIRONMENT
FOR ACADEMIC RESEARCH

GitLab

slurm
workload manager

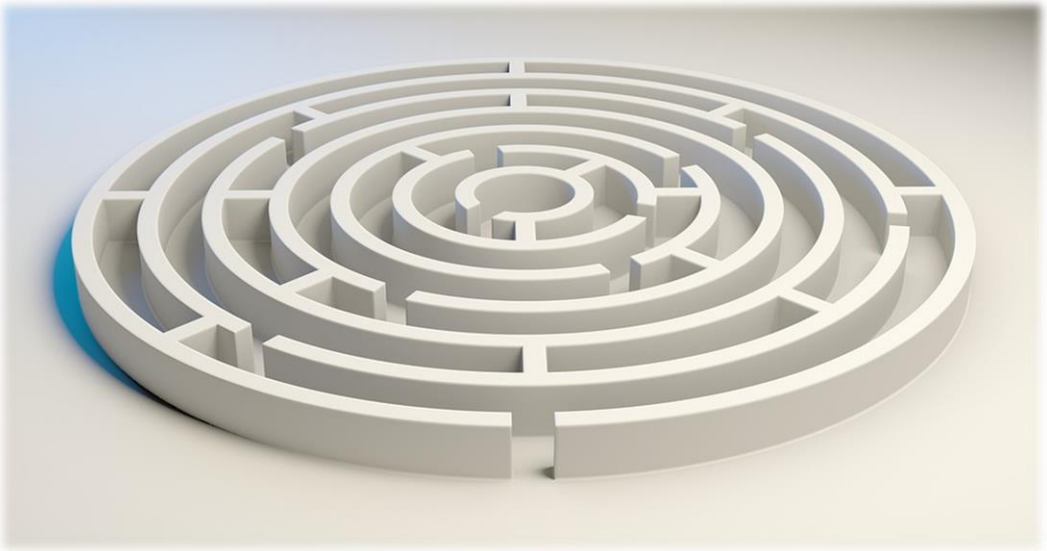
BEAR DataShare

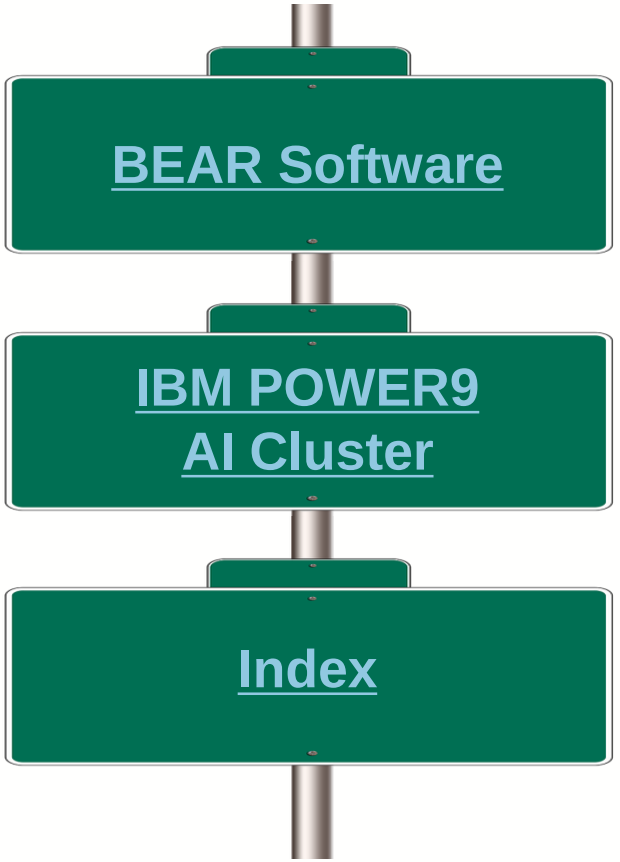


IBM® POWER9™

UNIVERSITY OF BIRMINGHAM | RESEARCH SOFTWARE GROUP

Questions?




BEAR Software
IBM POWER9
AI Cluster
Index

RSEConUK 2019

17th-19th September

University of Birmingham

- @RSEConUK
- https://rse.ac.uk/conf2019

"The Fourth RSE Conference will open on Tuesday 17th September and run for three full days, closing on Thursday 19th September 2019. We are not only expanding from two days to three but also opening up more delegate places to a maximum of 420 tickets."

RSEConUK 2019

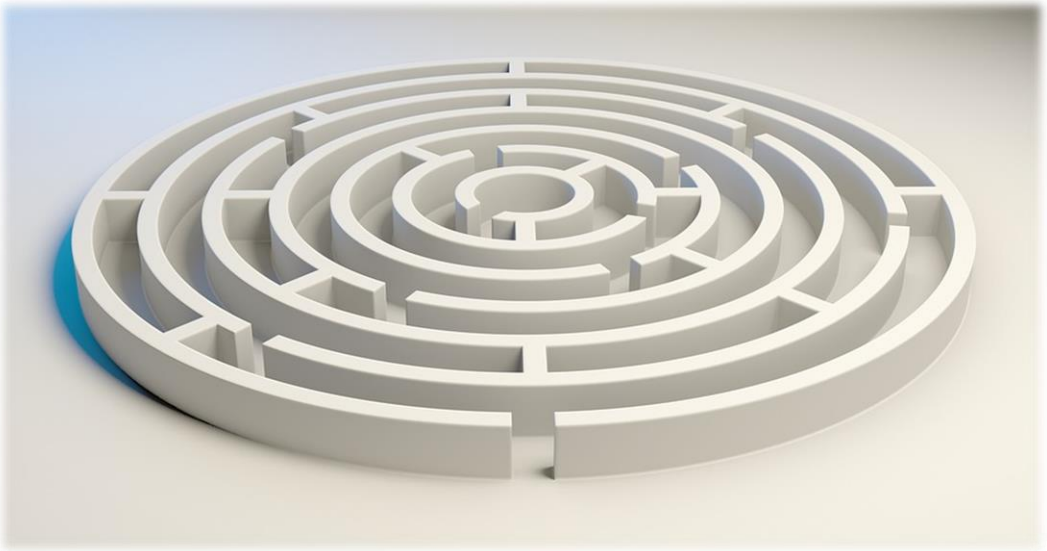
17th-19th September

University of Birmingham

Key Dates:

7th May	Call for Participation closes Call for Volunteers opens
23-30 May	Notification of successful submissions
31st May	Call for Volunteers closes
8th June	Detailed programme published on the website Successful volunteers notified
17th June	Registration opens
13th August	Registration closes

Questions?



University of Birmingham

Manuscripts and Phylogenetics

Index

Institute for Textual Scholarship and Electronic Editing

“An Analysis of the Coherence-Based Genealogical Method using Phylogenetics”

UNIVERSITY OF BIRMINGHAM | RESEARCH SOFTWARE GROUP

inside HPC

News | HPC Hardware | HPC Software | Industry Segments | White Papers | Resources | Special Rep

Sign up for our newsletter and get the latest HPC news and analysis. Email Address

Home » Archives for University of Birmingham

HPC Bear Cloud to Power Research at University of Birmingham

October 18, 2016 by staff | Leave a Comment

Designed specifically with researchers in mind, the Birmingham Environment for Academic Research (BEAR) Cloud will augment an already rich set of IT servic[illegible]y of Birmingham and will be used by ac[illegible]ics across all disciplines, [illegible]m Medicine to Archaeol[illegible]y, and Physics to Theology. "We[illegible]e very proud of the new syste[illegible]but building a research cl[illegible]isn't easy," said Simon Thompson, [illegible]ing Infrastructure Architect in IT Services at the University of Birmingham. "We challenged a range of carefully-selected partners to provide the underlying technology."

Filed Under: Cloud HPC, Compute, Datacenter, HPC Hardware, HPC Software, Industry Segments, Network, New Installations, News, Research / Education, Storage Tagged With: BEAR cloud, DDN, IBM, IBM Spectrum Scale, lenovo, Mellanox, University of Birmingham

"Non-traditional use of HPC"

UNIVERSITY OF BIRMINGHAM

Teaching excellence | Alumni | Work here | Events | Visit | UK | Dubai

Study | Research | International | Business | News | About us

Home > News and events > News > New high-performance computing system launched to support research at University

New high-performance computing system launched to support research at University of Birmingham

Posted on 17 Oct 2016

Research at the University of Birmingham is set to be given a boost this week with the launch of a high-performance, bespoke computing system known as BEAR Cloud.

Designed specifically with researchers in mind, the Birmingham Environment for A[illegible]h (BEAR) Cloud is far more powerful than any individual computer and can perform b[illegible]ns of calculations a sec[illegible]. It will be used by academics across all disciplines, from Medicine to Archaeolog[illegible]and Physics to Theology.

The system, developed by the IT Services department at the University of Birmingham, [illegible] already been used in the fight against deadly diseases such as Zika and Ebola, and to improve our understanding and treatment of diseases such as Cancer.

TABOR NETWORK: DATANAMI | ENTERPRISETECH | HPCWIRE JAPAN | LEVERAGE BIG DATA + ENTERPRISE HPC

HPC wire

Since 1987 - Covering the Fastest Computers in the World and the People Who Run Them

- Home
- Technologies
- Sectors
- Exascale
- Specials
- Resource Library
- Events
- Job Bank
- About

New HPC System Launched to Support Research at University of Birmingham

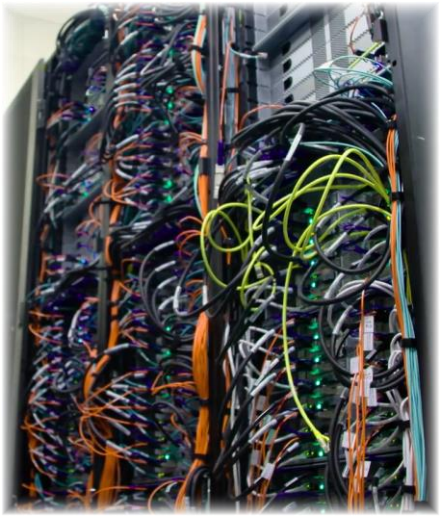
October 18, 2016

Oct. 18 — Research at the University of Birmingham is set to be given a boost this week with the launch of a high-performance, bespoke computing system known as BEAR Cloud.

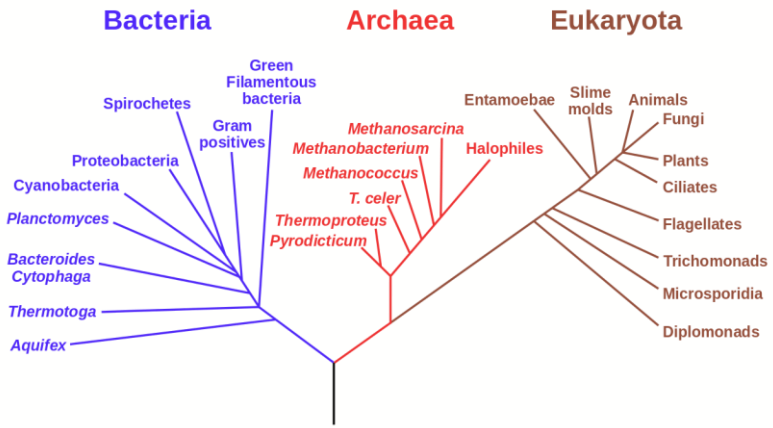
Designed specifically with researchers in mind, the Birmingham Environment for Academic Research (BEAR) Cloud will augment an [illegible] IT services at the University of Birmingham and will be [illegible]ed by academics across all disciplines, from Medicine to Archaeolog[illegible]and Physics to Theology.

BEAR Cloud retains the power of traditional High Perfo[illegible]ce Computin[illegible] (HPC) while benefitting from the latest virtualisation technology built on OpenStack. OpenStack technology is used to deliver more flexibility and ease of access than is found in traditional HPC batch processing environments. This private cloud facility will provide an environment in which it is easy to deploy and reconfigure instances on demand to meet the computationally intensive needs of research groups. Demands will be highly variable but will include HPC clusters, dedicated workflows such as Galaxy for genomics, and Matlab processing farms.

An Analysis of the Coherence-Based Genealogical Method using Phylogenetics



New Testament Greek Manuscripts

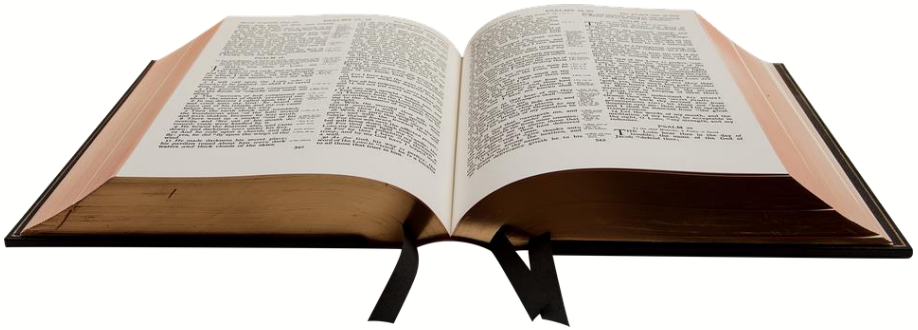


The Coherence-Based Genealogical Method (CBGM)

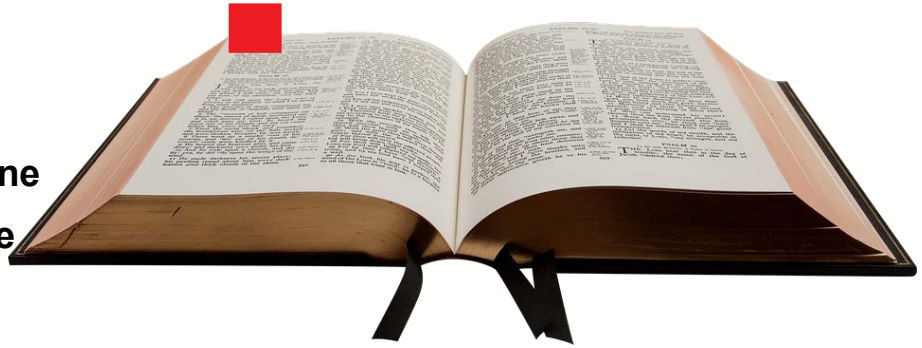
(What's it for?)

The search for the original text of the Bible (Traditional method)

A few manuscripts



The search for the original text of the Bible (20th Century Problem)



Over 1,500 known manuscripts of John's Gospel alone

Over 6,500 places of variation in John's Gospel alone

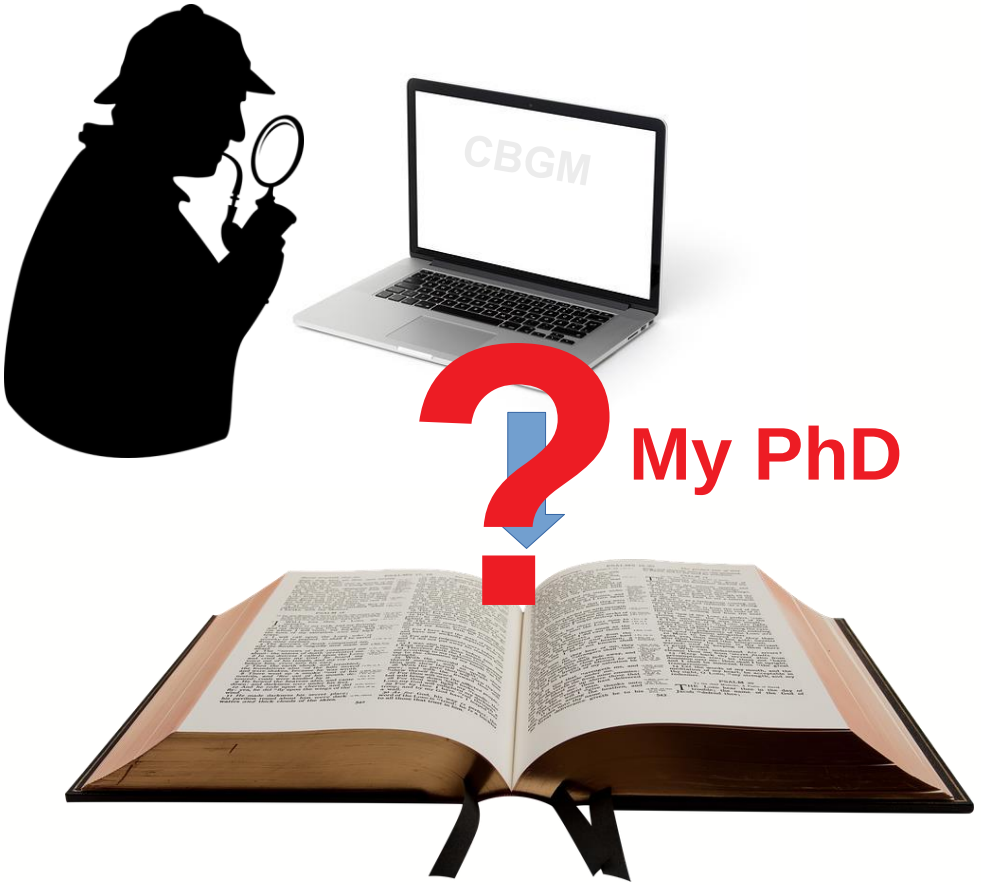
-1 Verse 21 +1

Then they were willing to take him into the boat, and immediately the boat reached the shore where they were heading

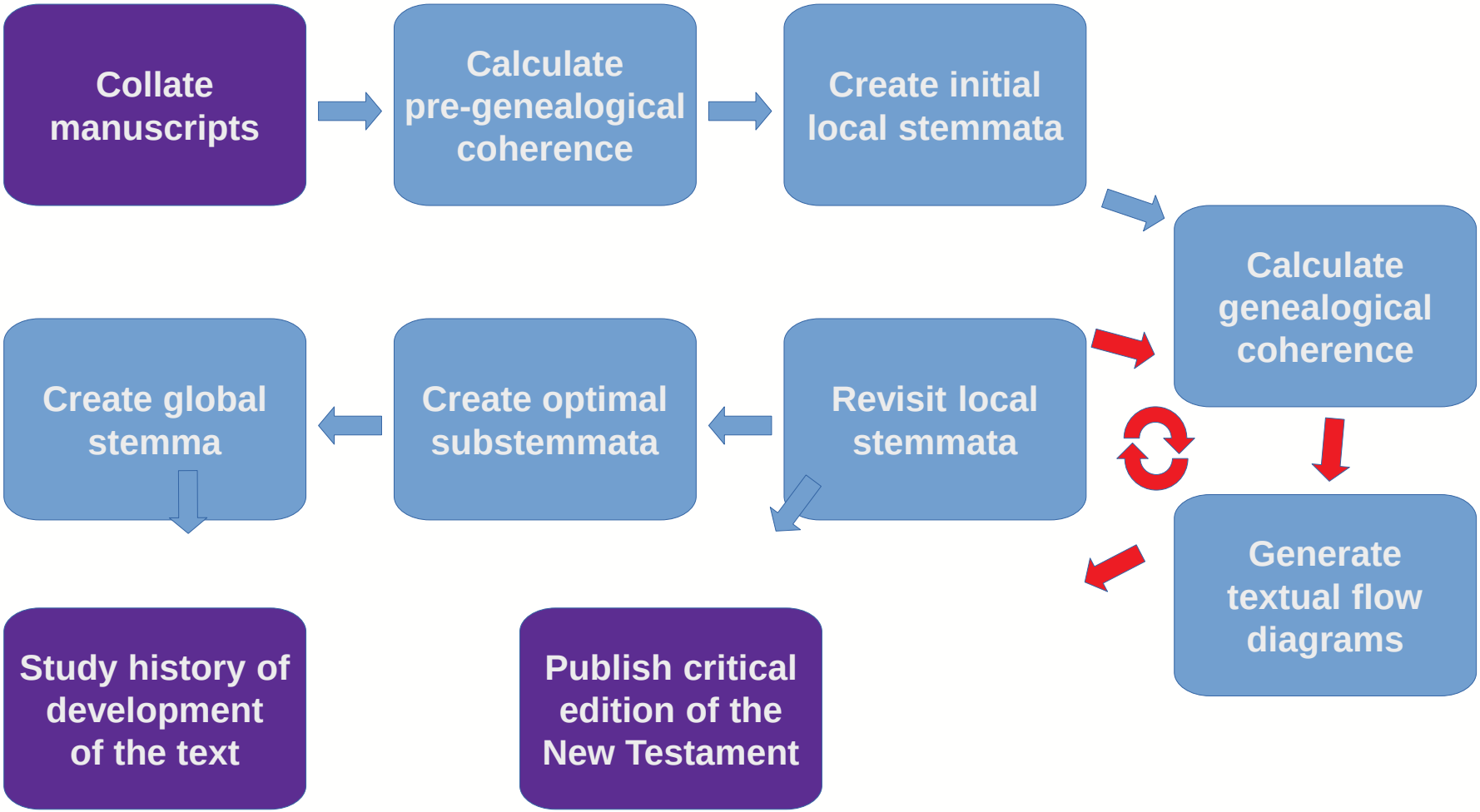
01	100%	ηλθον ουν λαβιν αυτον εις το πλοιον και ευθεως το πλοιον εγενετο επι την γην εις ην υπηντησεν
157 1010 1293 1424 1561 L329 L640	94.1%	ηθελον ουν λαβειν αυτον εις το πλοιον και ευθεως το πλοιον εγενετο επι την γην εις ην υπηγον
01 *correctorca*	94.0%	ηλθον ουν λαβιν αυτον εις το πλοιον και ευθεως το πλοιον εγενετο επι της γης εις ην υπηγον
L32	93.0%	ηθελον ουν λαβειν αυτον εις το πλοιον και ευθεως το πλοιον εγενετο εις την γην εις ην υπηγον
0211	93.0%	ηθελον ουν λαβειν αυτον εις το πλοιον και ευθεως το πλοιον εγενετο επι την γην εις ην υπιγον
L847	93.0%	ηθελον ουν λαβειν αυτον εις το πλοιον και ευθεως το πλοιον εγενετο επι την γην εις ην υπιγων
TR 𝔓28 07 09 013 017 028 030 031 034 036 037 039 *corrector1* 045 047 18 22 35 168 226 265 295 544 841 992 1009 1029 1079 1128 1192 1210 1219 1230 1278 1320 1344 1463 *corrector* 1546 1571S 1797 2372 2411 2615 2718 2766 *corrector* 2786 2790 L638 L663 L770 L1075 L1086 L1091 L1096 L1552	91.9%	ηθελον ουν λαβειν αυτον εις το πλοιον και ευθεως το πλοιον εγενετο επι της γης εις ην υπηγον
2585 L141	91.9%	ηθελον ουν λαβειν αυτον εις το πλοιον και ευθεως το πλοιον εγενετο επι της γης εις ην υπηγων
021	90.8%	ηθελον ουν λαβειν αυτον εις το πλοιον και ευθεος το πλοιον εγενετο επι της γης εις ην υπηγον
1014	90.8%	ηθελον ουν λαβειν αυτον εις το πλειον και ευθεως το πλοιον εγενετο επι της γης εις ην υπηγον
L425 *corrector*	90.8%	ηθελον ουν λαβειν αυτον εις το πλοιον και ευθεως το πλοιον εγενετο επι της γης εις ην υπιγων
L5	90.8%	ηθελον ουν λαβειν αυτον εις το πλοιον και ευθεως τω πλοιον εγενετο επι της γης εις ην υπηγον
038	90.8%	ηθελον ουν λαβειν αυτον εις το πλοιον και ευθεως το πλοιον εγενετο επι της γης εις ιν υπηγον
039	90.7%	ηθελον ουν λαβειν αυτον εις το πλοιον και ευθεως το πλοιον εγενε επι της γης εις ην υπηγον
1253	90.6%	ηθελον ουν λαβειν αυτον εις το πλοιον και ευθεως το πλοιον εγενετο εις την γην ην υπηγον
173	90.2%	ηθελον ουν λαβειν αυτον εις το πλοιον και ευθεως το πλοιον εγενετο επι της γης ης ην υπηγων
L252	89.9%	ηθελον ουν λαβειν αυτον εις το πλοιον και ευθεως ουν το πλοιον εγενετο επι της γης εις ην υπηγον
L704 L1073 L1076	89.8%	ηθελησαν ουν λαβειν αυτον εις το πλοιον και ευθεως το πλοιον εγενετο επι της γης εις ην υπηγον
L1000	89.7%	ηθελον ουν λαβειν αυτον εις τω πλοιον και ευθεως το πλοιον εγενετο επι της γης εις ην υπιγον
L425	89.7%	ηθελον ουν λαβειν αυτων εις το πλοιον και ευθεως το πλοιον εγενετο επι της γης εις ην υπιγων
1319	89.7%	ηθελον ουν λαβειν αυτον εις το πλοιον και ευθεως το πλοιον εγενετο δια της γης εις ην υπηγον
L387	89.7%	ηθελον ουν λαβειν αυτον εις το πλοιον και ευθεως το πλοιον εγενετο επι της γης εις ιν υπιγον
2106	89.5%	ηθελον λαβειν αυτον εις το πλοιον και ευθεως το πλοιον εγενετο επι της γης εις ην υπηγον
L735	89.1%	ηθελων ουν λαβην αυτον εις το πλοιον και ευθεως τω πλοιον εγενετο επι της γης εις ην υπηγων
131 872	88.6%	ηθελον ουν λαβειν αυτον εισ το πλοιον και ευθεωσ το πλοιον εγενετο επι τησ γησ εισ ην υπηγον
732 1463	88.3%	ηθελον ουν λαβειν εις το πλοιον και ευθεως το πλοιον εγενετο επι της γης εις ην υπηγον
377	88.2%	ηθελον λαβειν αυτον εισ το πλοιον και ευθεωσ το πλοιον εγενετο επι τησ γησ εισ ην υπηγον

The search for the original text of the Bible (21st Century Method)

The search for the original text of the Bible (21st Century Method)



CBGM Process

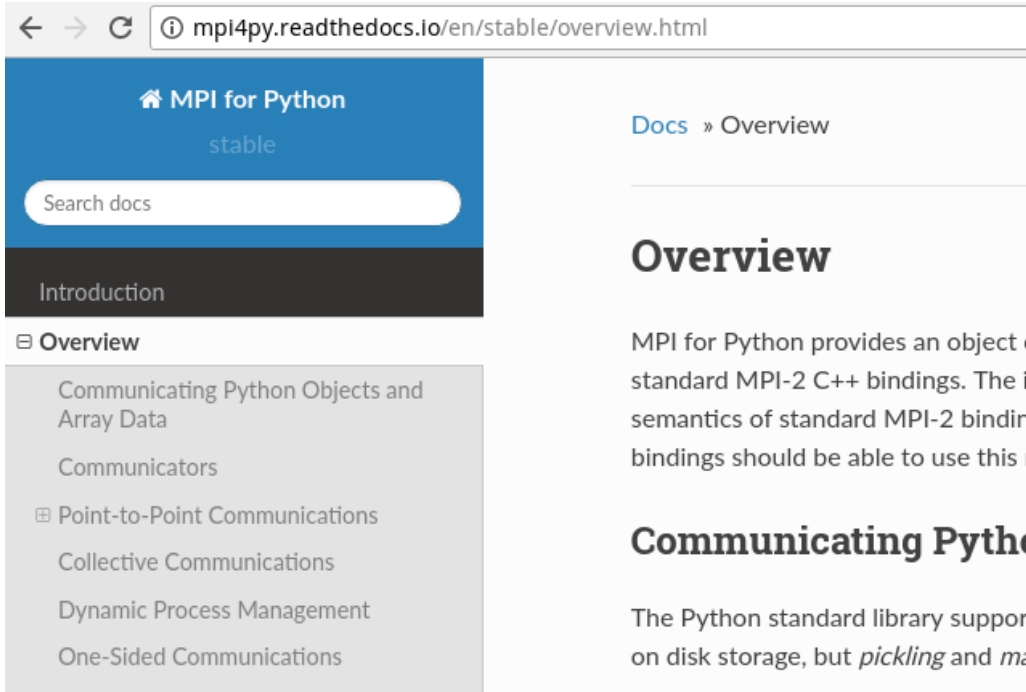


Implementing the CBGM

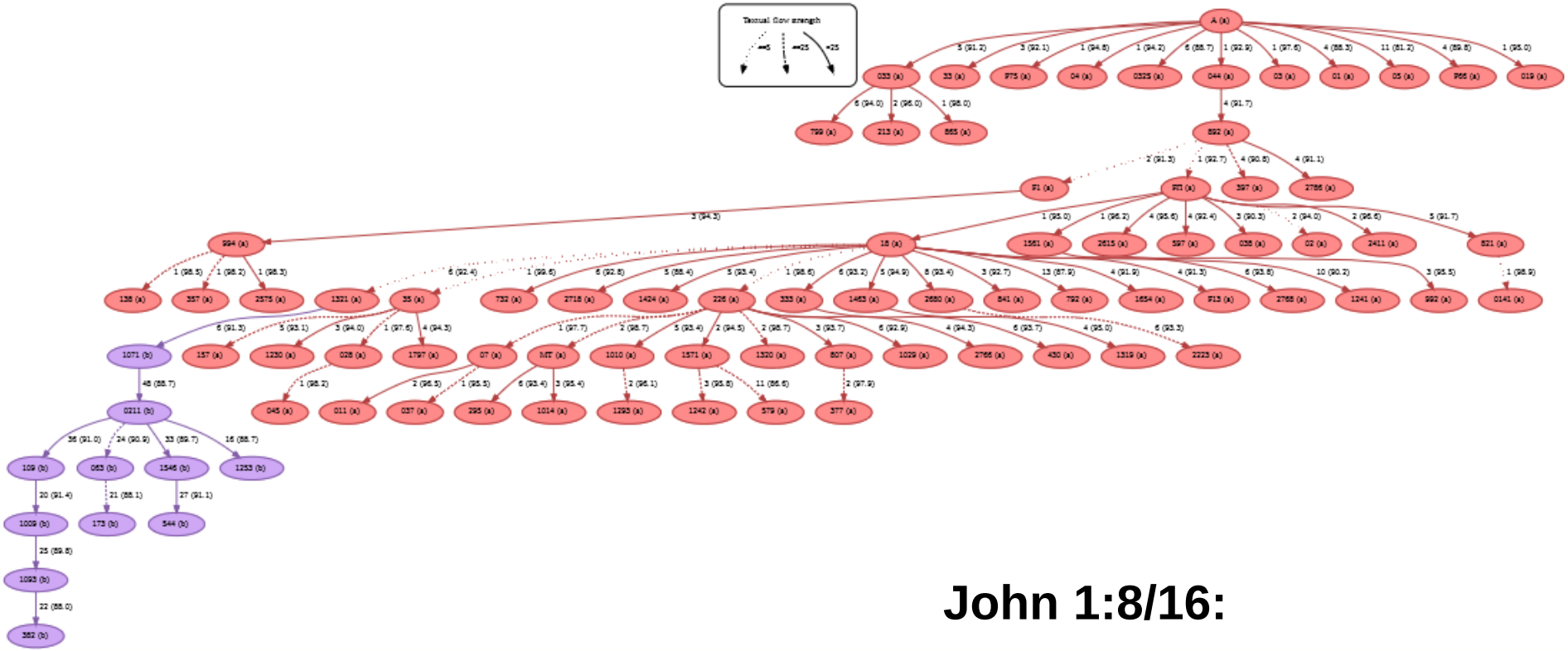
BEAR

BIRMINGHAM ENVIRONMENT FOR ACADEMIC RESEARCH

UNIVERSITY OF BIRMINGHAM | RESEARCH SOFTWARE GROUP



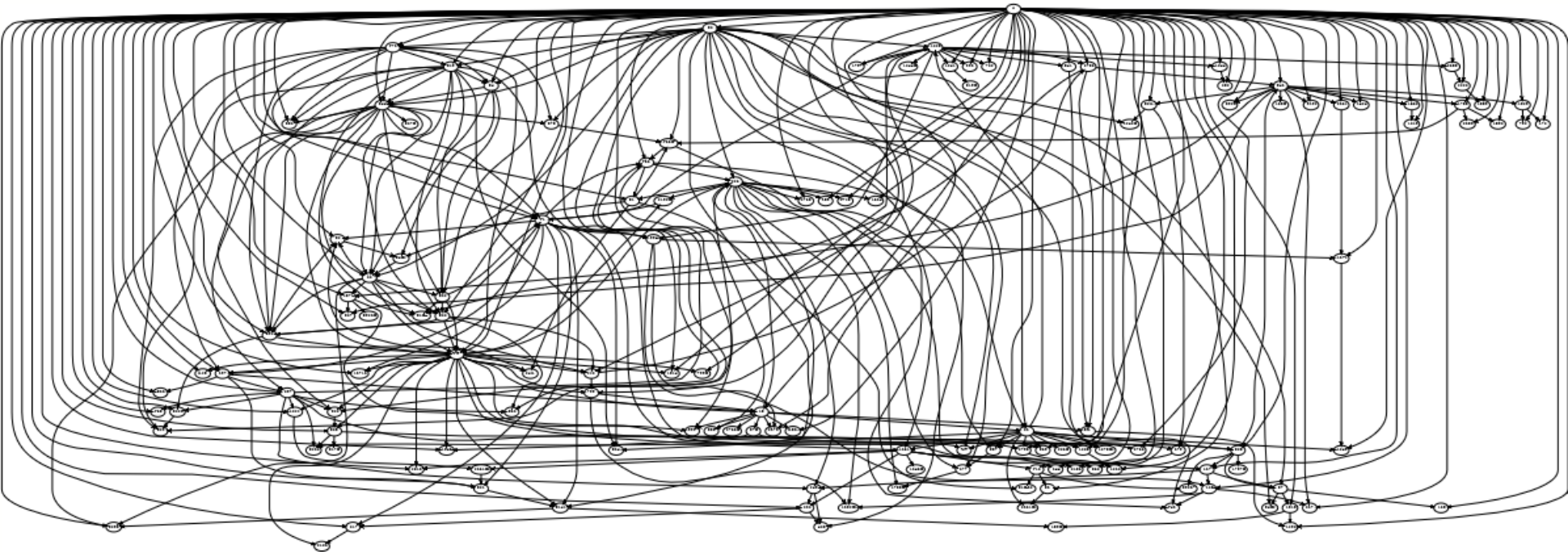
Textual flow diagram



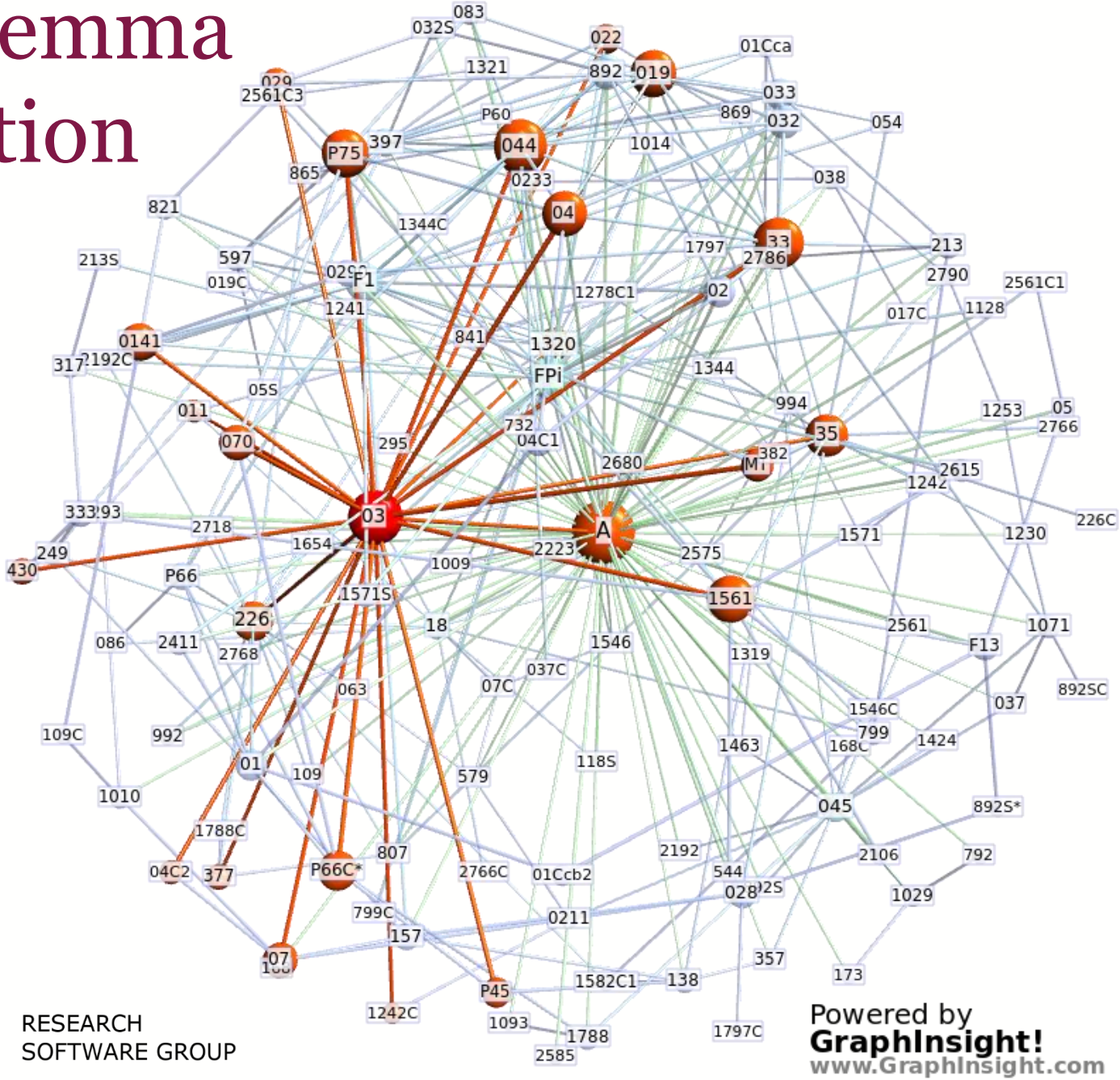
John 1:8/16:

μαρτυρηση vs. μαρτυρησει

Global stemma



Global stemma visualisation



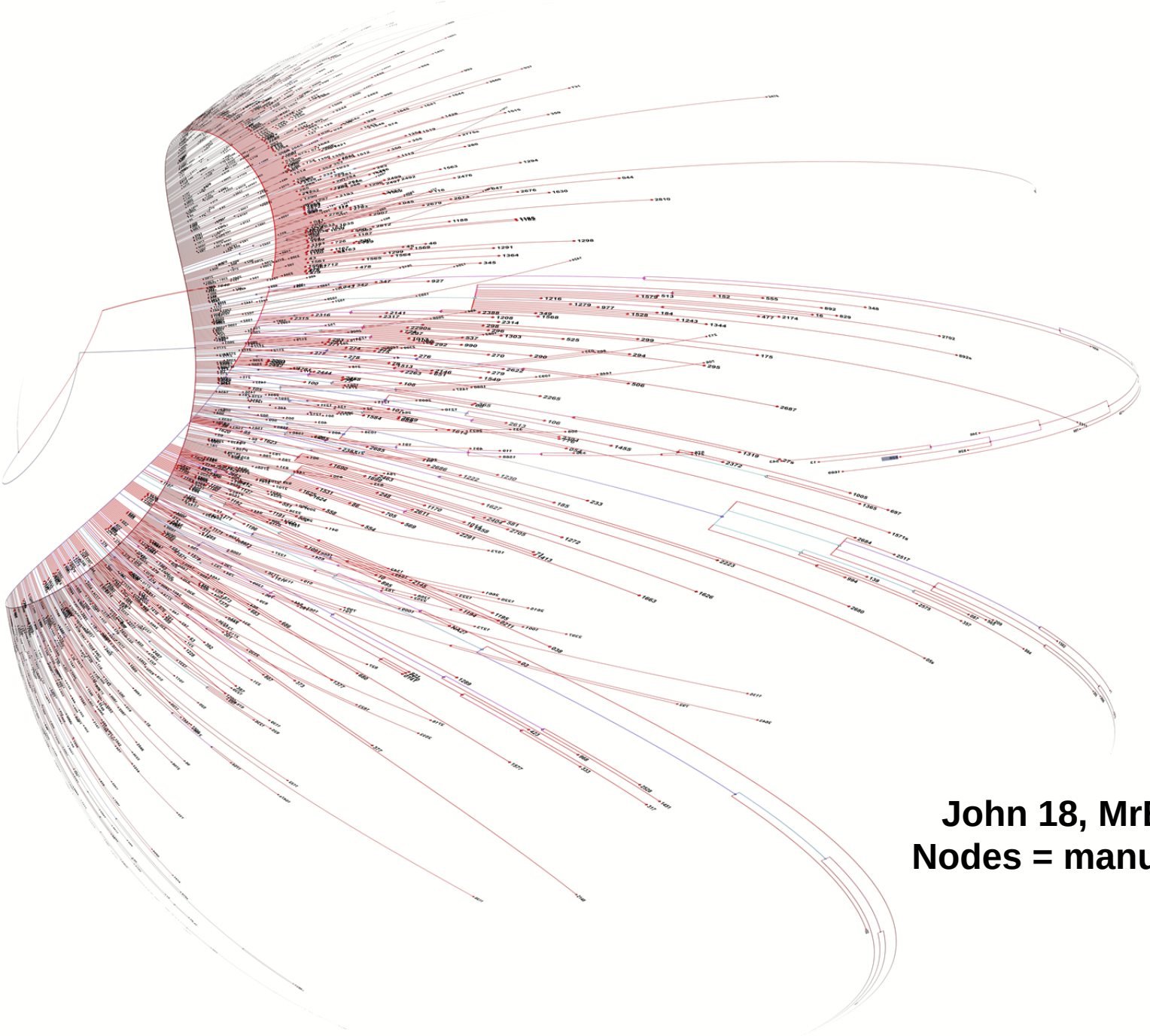
UNIVERSITY OF BIRMINGHAM | RESEARCH SOFTWARE GROUP

Powered by **GraphInsight!**
www.GraphInsight.com

Phylogenetics for the Bible (In three slides)

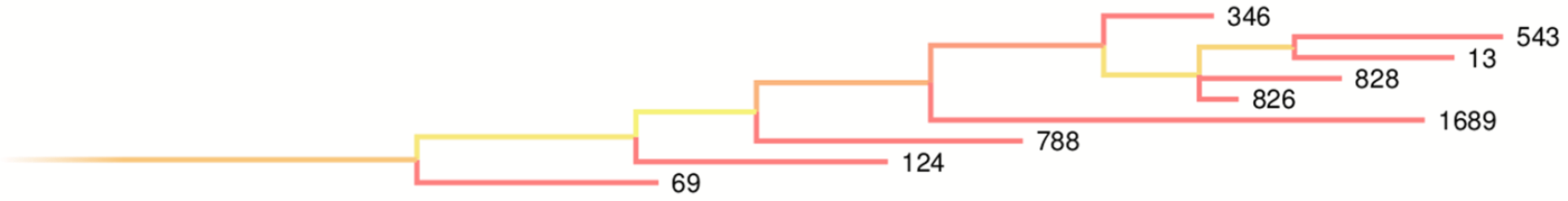
Phylogenetics for the Bible

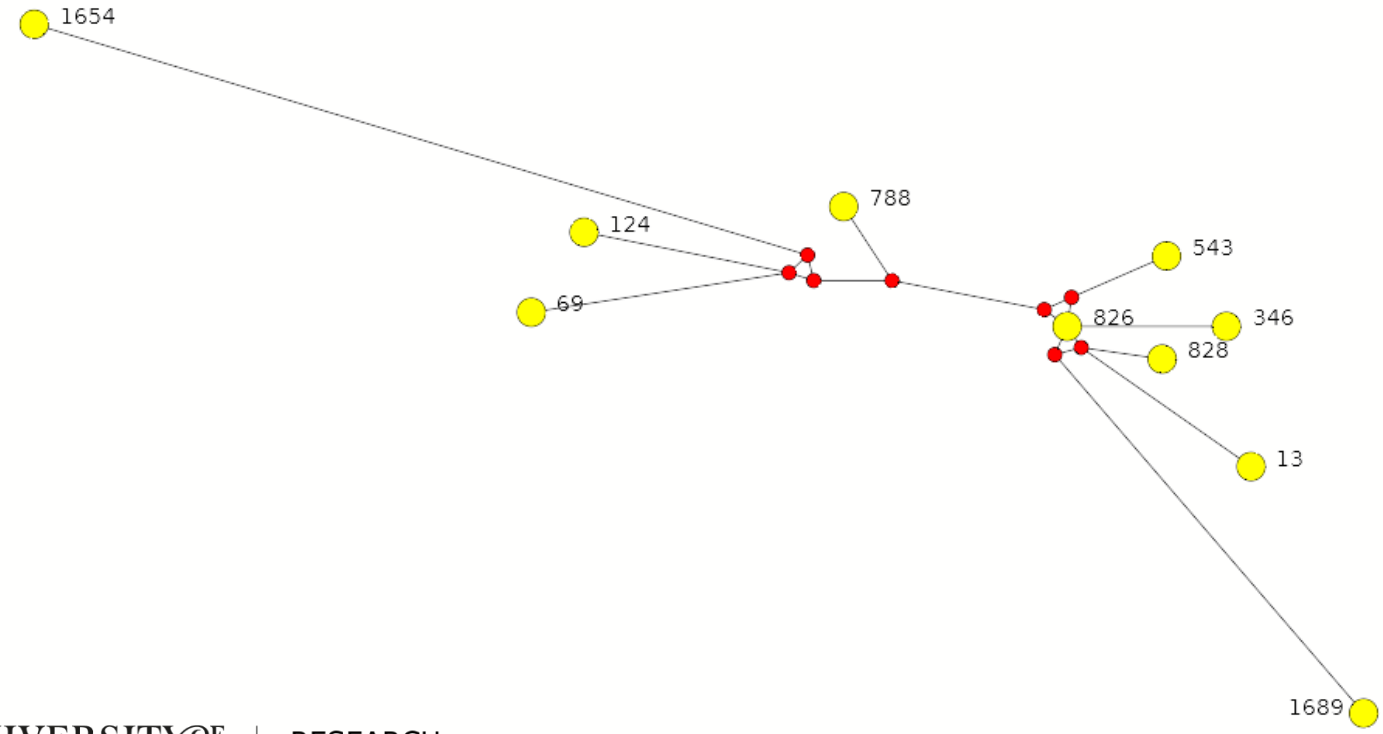
Encode



John 18, MrBayes
Nodes = manuscripts

Family 13 - Phylogenetics



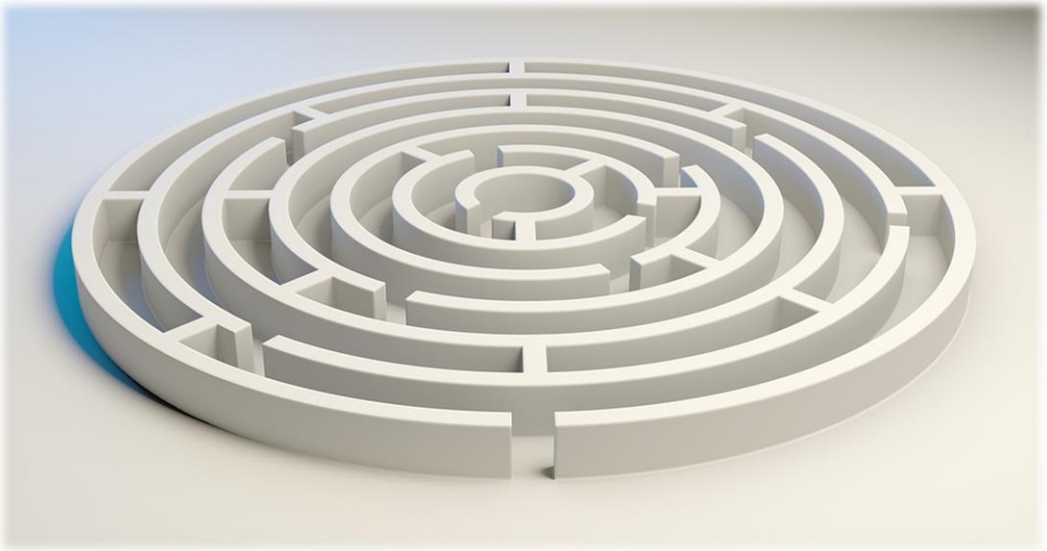


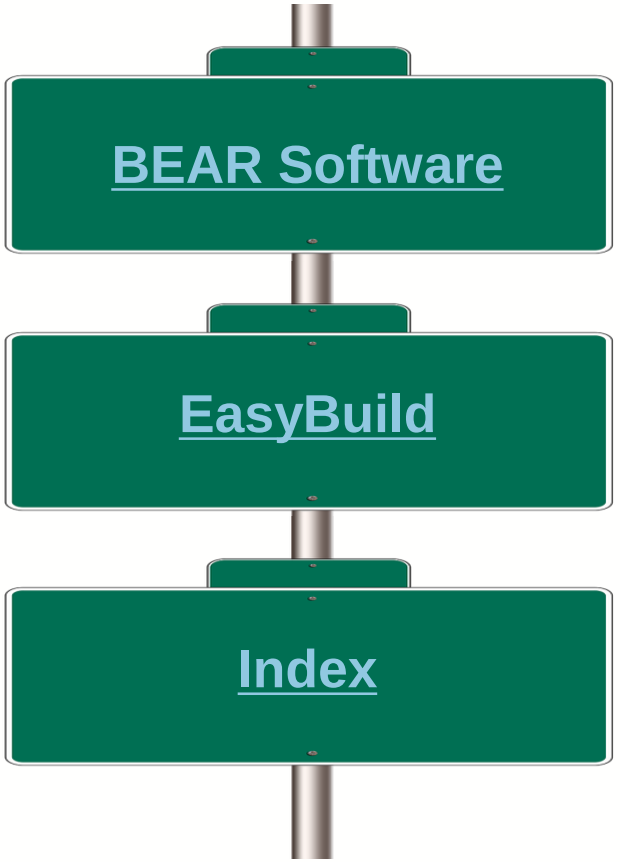
PhD Conclusions?

- Phylogenetics works very well with this kind of data
- Phylogenetic and CBGM experiments effectively agree – and agree with traditional scholarship
- => Both methods basically sound
- Lots of recommendations for small improvements to the CBGM

Manuscripts and Phylogenetics

Questions?




BEAR Software
EasyBuild
Index

BEAR Software

BEAR Software – GOALS

- Enable the University of Birmingham's research community to get the best from their research software: "Better software, better research"
- Provide specialist software engineering advice and support to researchers and RSEs
- Help to enhance the University's reputation for high quality research
- Help researchers get the most from BEAR Services, maximising the return on the University's investment in BEAR.

Advice

- Examples:
  - How best to use HPC to analyse data from Cyclotron experiments
  - Transitioning to using git and GitLab
  - Parallel MATLAB use
  - How to structure Django applications
  - Use of HPC VM for health sciences

UNIVERSITY OF BIRMINGHAM | RESEARCH SOFTWARE GROUP

Coding

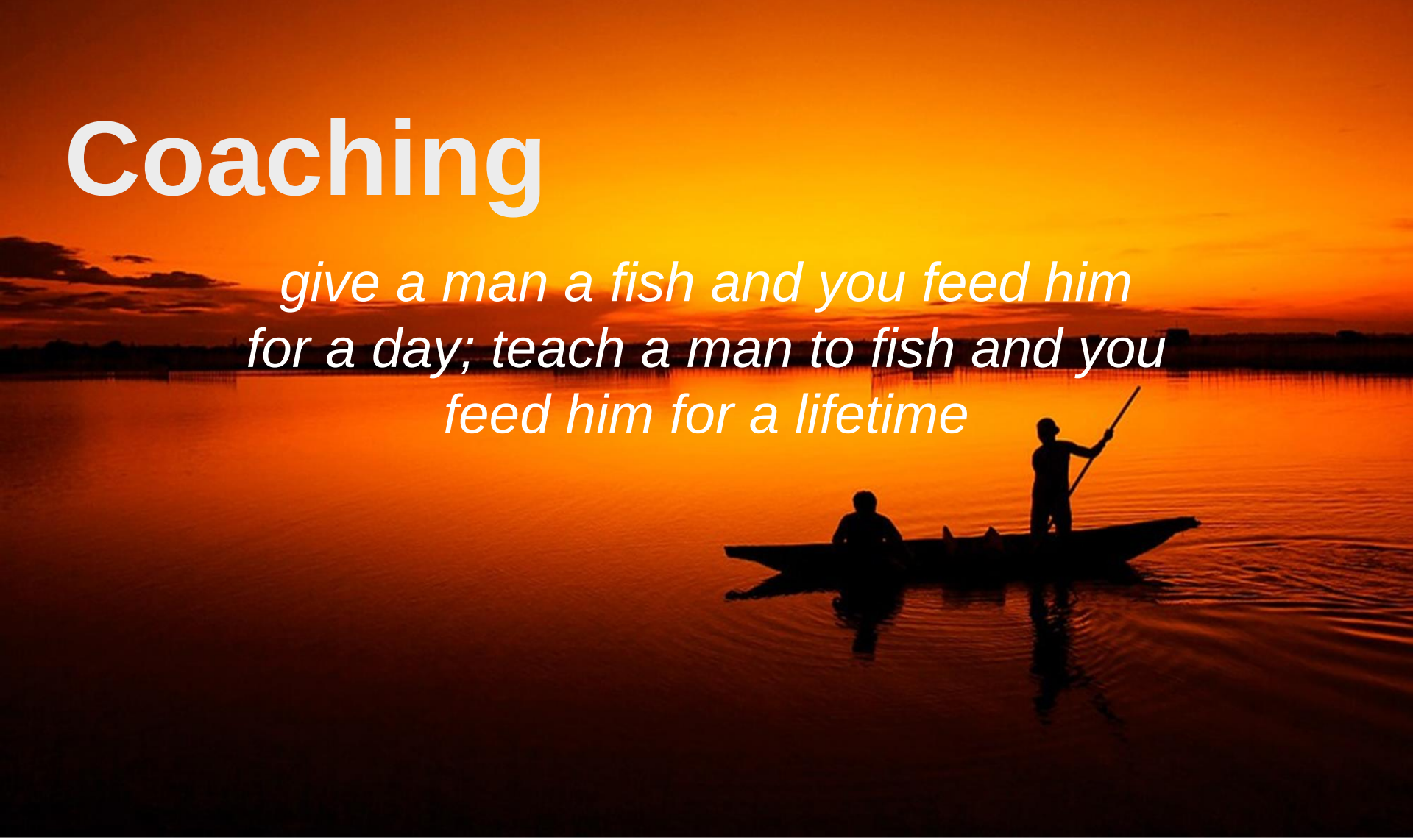
- Up to 10 days of free coding. Examples:
  - Replace outdated web application
  - Migrate MPI library to C++
  - Create an application to track the recycling process of batteries
  - Parallelise R script to run on HPC

UNIVERSITY OF BIRMINGHAM | RESEARCH SOFTWARE GROUP

Coding

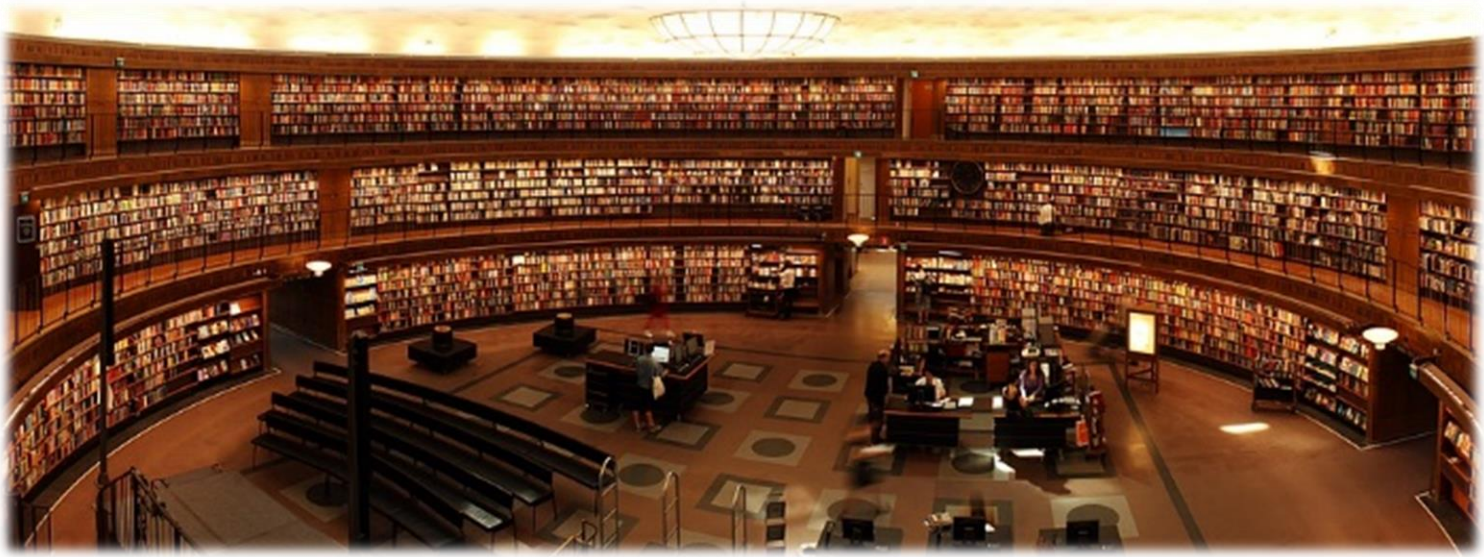
10 RSEs » **Max 20 projects / month @ 10 days / project**

Coaching


give a man a fish and you feed him for a day; teach a man to fish and you feed him for a lifetime

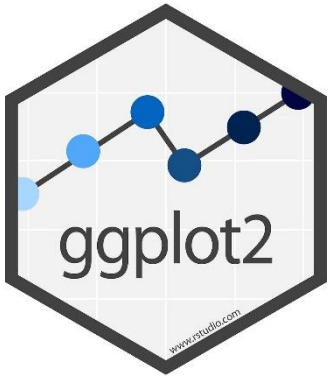
Coaching

10 RSEs » **Impacting 100 projects / month @ 1 session / week**



Coaching

- Up to 20 half-day sessions of free coaching
- Examples:
  - Developing a web application
  - Coaching a department RSE
  - Developing a mobile app
  - Geospatial analysis in R
  - Experiment with C++ library



Fortran

UNIVERSITY OF BIRMINGHAM | RESEARCH SOFTWARE GROUP

Training

- Software Carpentries
  - Python, R, Git, Bash

- NVIDIA Deep Learning

- Intro to HPC

DEEP LEARNING INSTITUTE

UNIVERSITY OF BIRMINGHAM | RESEARCH SOFTWARE GROUP

Annual report

UNIVERSITY OF BIRMINGHAM

BEAR
BIRMINGHAM ENVIRONMENT FOR ACADEMIC RESEARCH

RESEARCH SOFTWARE GROUP
BEAR - Advanced Research Computing

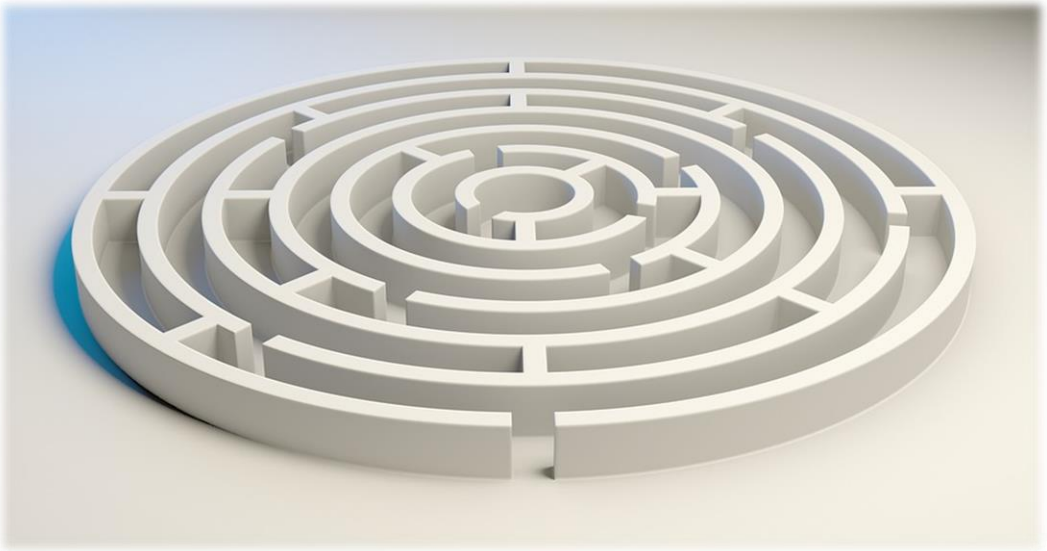
bear-software@contacts.bham.ac.uk | @uob_rescomp
www.birmingham.ac.uk/bear-software

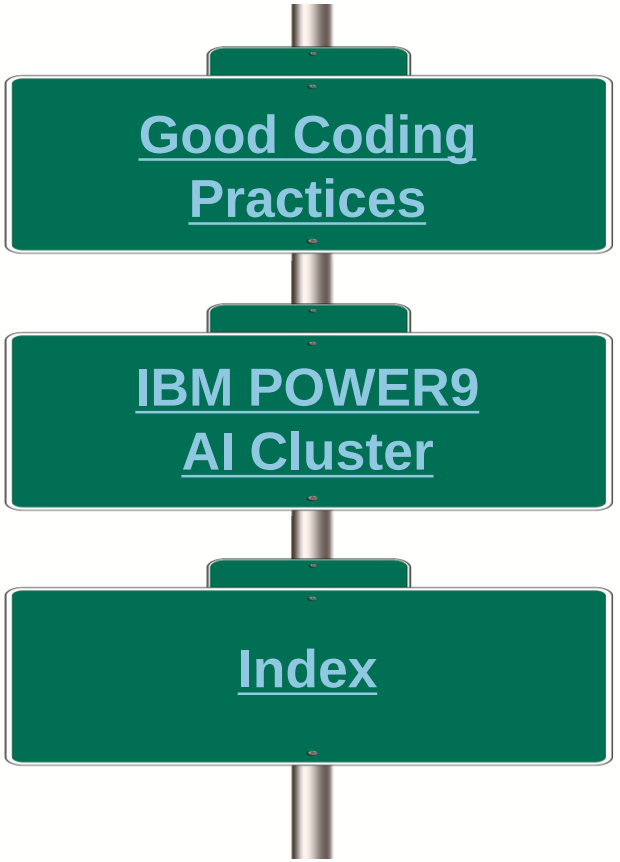
2018 Report

https://intranet.birmingham.ac.uk/it/teams/infrastructure/research/bear/documents/public/RSG-2018-Report-v4-approved.pdf

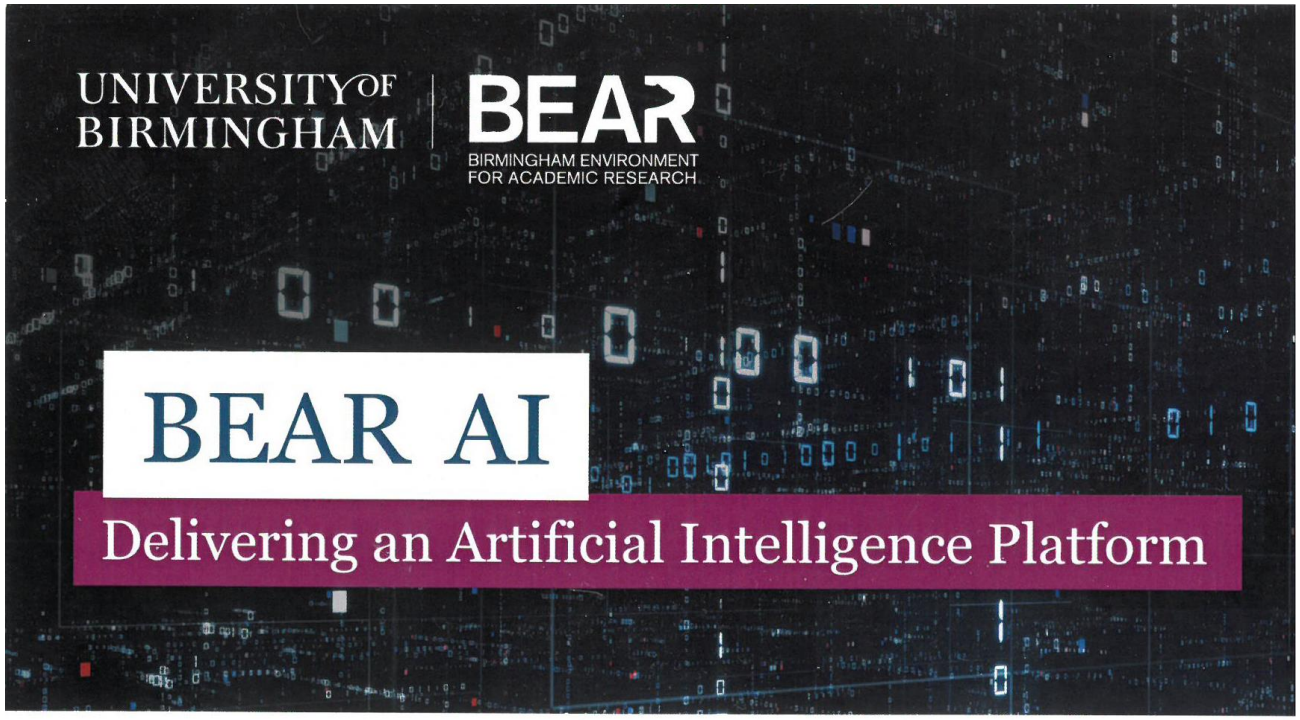
BEAR Software

- Questions?




Good Coding Practices
IBM POWER9 AI Cluster
Index

BEAR AI


UNIVERSITY OF BIRMINGHAM
BEAR
BIRMINGHAM ENVIRONMENT FOR ACADEMIC RESEARCH
BEAR AI
Delivering an Artificial Intelligence Platform
bearinfo@contacts.bham.ac.uk | @uob_rescomp
intranet.birmingham.ac.uk/bear-ai

IBM® POWER9™ AI Cluster

"Researchers at the University of Birmingham are set to benefit from the largest IBM® POWER9™ Artificial Intelligence (AI) computer cluster in the UK."

Currently 11 nodes, each with:

- Dual IBM POWER9 CPUs with 18 cores each, which currently present themselves as 144 cores using simultaneous multithreading (SMT4).
- Four NVIDIA Tesla V100 Tensor Core GPUs
- 1 TB system memory
- High speed NVIDIA NVLink interconnect fully meshed between the GPUs and also into the system memory
- 100G InfiniBand interconnect to other nodes and storage systems

https://intranet.birmingham.ac.uk/bear-ai

UNIVERSITY OF BIRMINGHAM | RESEARCH SOFTWARE GROUP

GPU-accelerated Software

Installed version: 1.0.1
An open source deep learning platform from Facebook.

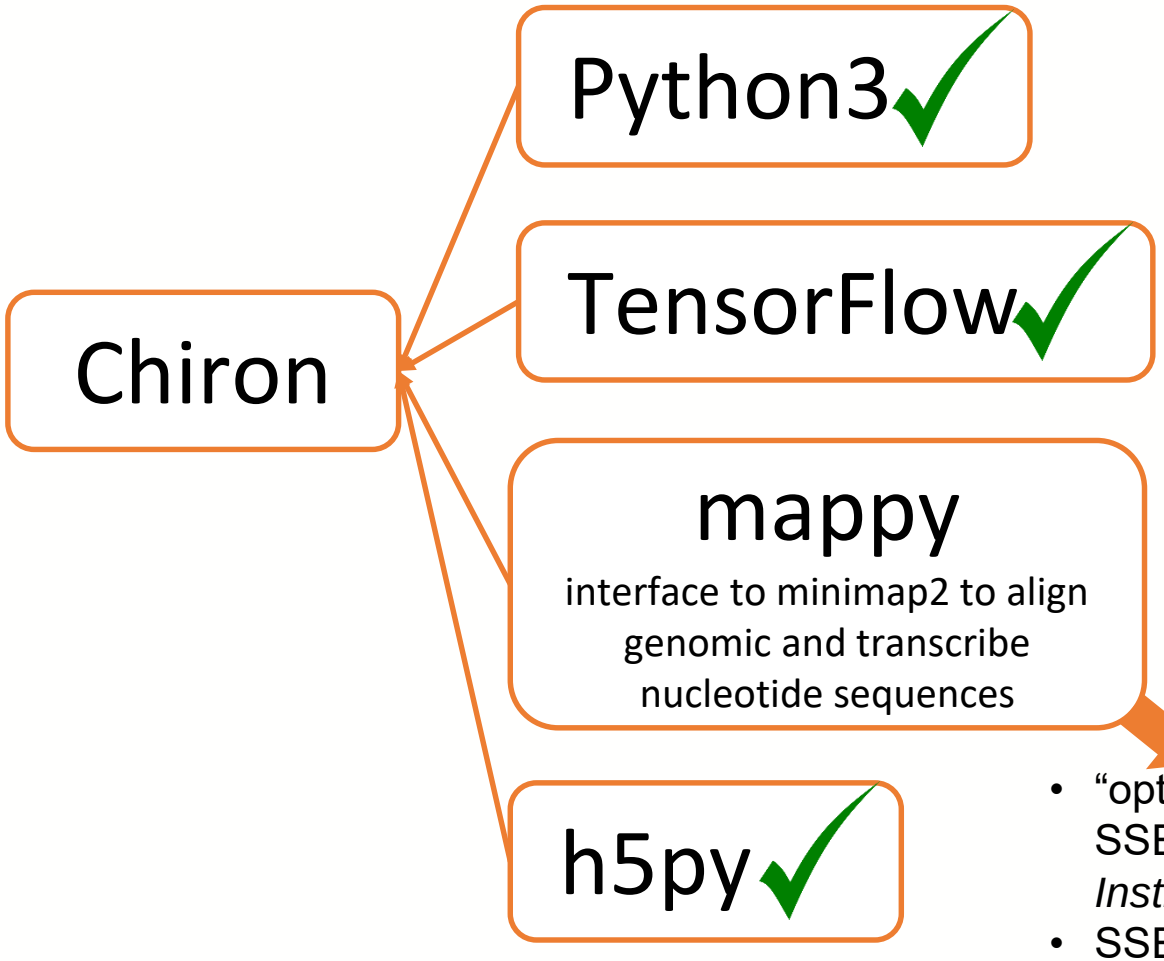
Installed version: 1.10.1
Python-based open source machine learning framework from Google.

Amber

Installed version: 18
A suite of biomolecular simulation programs with GPU acceleration

Installed version: 2018.4
HPC molecular dynamics package with GPU acceleration.

HPC Software



- "optimized for x86-64 CPUs" using SSE2: Intel extensions for "*Single Instruction, Multiple Data*"
- SSE2 code won't build on POWER
- POWER equivalent is "Altivec"
- So we need to port the C code…

Porting SSE to Altivec

- https://github.com/IvantheDugtrio/veclib implements a subset of the SSE2 -> Altivec conversion
- https://github.com/IvantheDugtrio/veclib/pull/6 is our PR which has been merged upstream, including our implementation of a missing function and some fixes for others
- Problem: there is no direct Altivec equivalent for SSE2's `_mm_cvtsi32_si128` function – so we had to write one.
- Now we have installed veclib on our POWER9 boxes – using EasyBuild – for whenever we need it.

Chiron ✓

Home About Join Events

MAY 7, 2019 BY ANDREW EDMONDSON

Second Power AI User Group Meeting

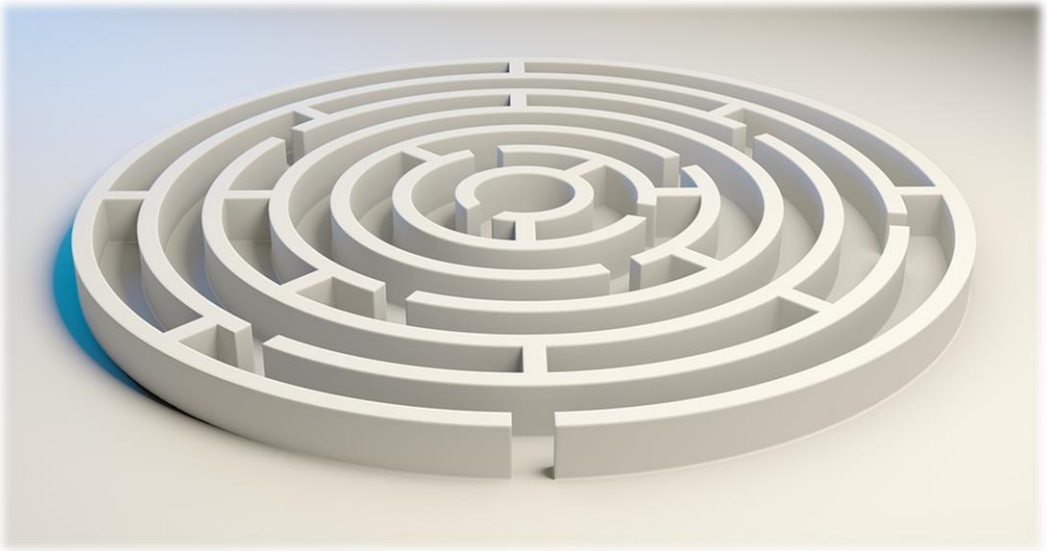
The second Power AI UG meeting will take place on Monday 8th July in Birmingham from 10am to 4pm. Register here.

https://www.poweraiug.org

UNIVERSITY OF BIRMINGHAM | RESEARCH SOFTWARE GROUP

BEAR AI

- Questions?




Good Coding Practices
EasyBuild
Index

EasyBuild

"EasyBuild is a software build and installation framework that allows you to manage (scientific) software on High Performance Computing (HPC) systems in an efficient way."

https://easybuilders.github.io/easybuild/

EasyBuild

- EasyBuild allows us to easily and reproducibly build software for various different platforms in BlueBEAR and BEAR Cloud.
- We have:
  - EL7 sandybridge, haswell, broadwell, skylake
  - Ubuntu 16.04 haswell
  - EL7 POWER9
  - And Cascade Lake has just arrived on site…

EasyBuild

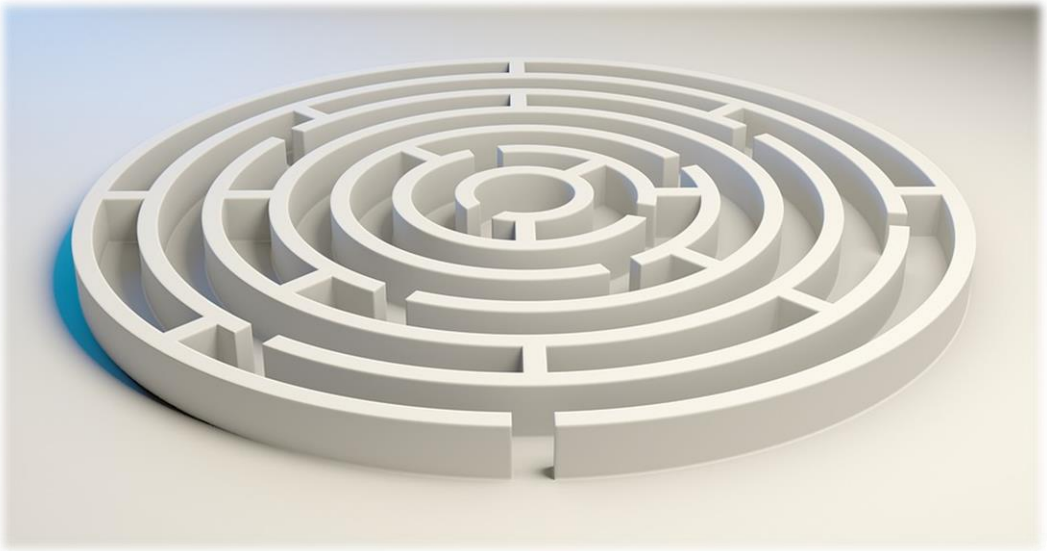
- Wasn't so "Easy" with POWER9
  - EasyBuild expects Intel
  - Lots of software doesn't build out of the box via EasyBuild on POWER9
- But… I've worked through various of the problems and am contributing it all upstream.

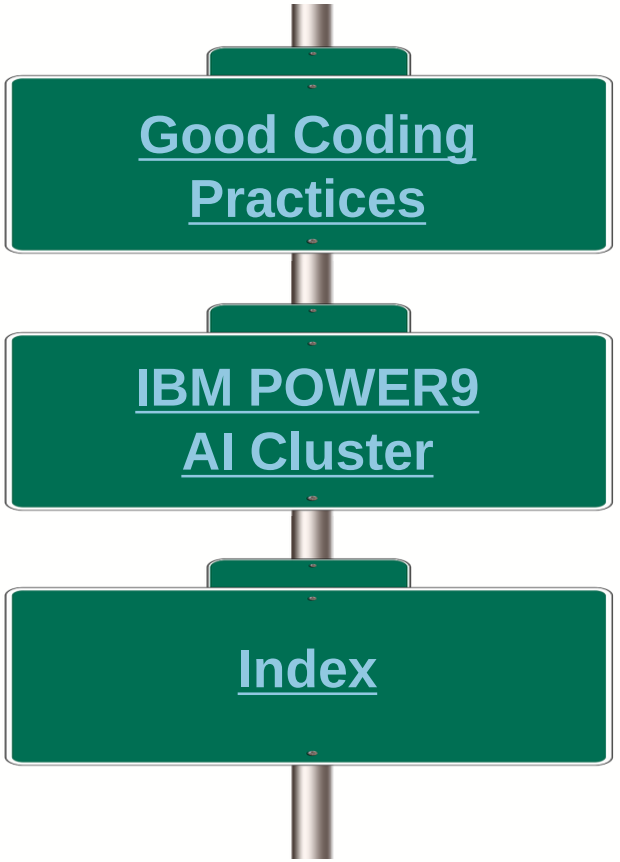
- And now it is "Easy"

EasyBuild

- Questions?

UNIVERSITY OF BIRMINGHAM | RESEARCH SOFTWARE GROUP




Good Coding Practices
IBM POWER9 AI Cluster
Index

Good Coding Practices

Reproducible Research

It's two years later… and you need to reproduce those results… but your software doesn't produce the same results today…

Reproducible Research

- **VERSION CONTROL**
  - Data
  - Software
  - Environment
- Does the same version of the software using the same version of the data running in the same environment produce the same results?

Reproducible Research

- If your software is not under control, then you cannot have confidence in your results. Or hope to reproduce them.

(or hope that anyone else could reproduce them either)

Chaos

Collaboration

Time has passed, you are now a PI. And you've got some RFs working with you now…

And they need to work on your code.

Collaboration (1/3)

- Version control (again!)
  - Branching, merging etc.
- Issue tracking
- Testing
  - Unit tests with continuous integration
- Building
  - Straightforward documentation
  - Use standard tools

Untested code

Collaboration (2/3)

- Comments – not too much, not too little
- Naming conventions
  - Use meaningful names
- Coding standards
  - E.g. PEP8 for Python
  - Static code analysis (while editing if possible)
- Make it readable

Misleading names / comments

Collaboration (3/3)

- Structure
  - Functions should do one thing
  - Functions should fit on one page
  - Files should contain related things
  - Folders should contain related things
- Simple is better than complex
  - Don’t write code you don’t need

Complexity

Editing your own code

It's 5 years later and you need to edit you own code… but you have no idea how it works.

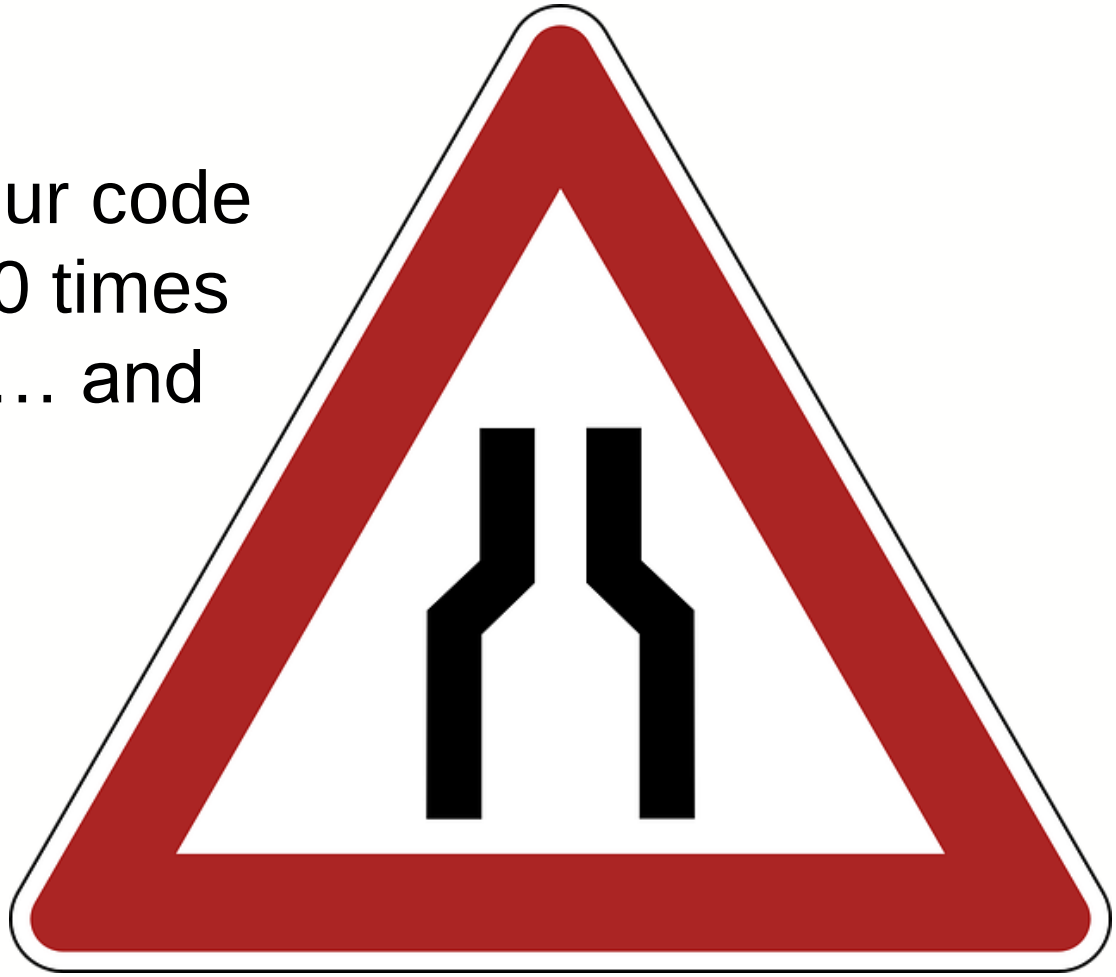
Editing your own code

- See “Collaboration”

Thinking “I’ll never need to…”

Optimisation

You need to run your code with data 1,000,000 times bigger than before… and it's really slow.



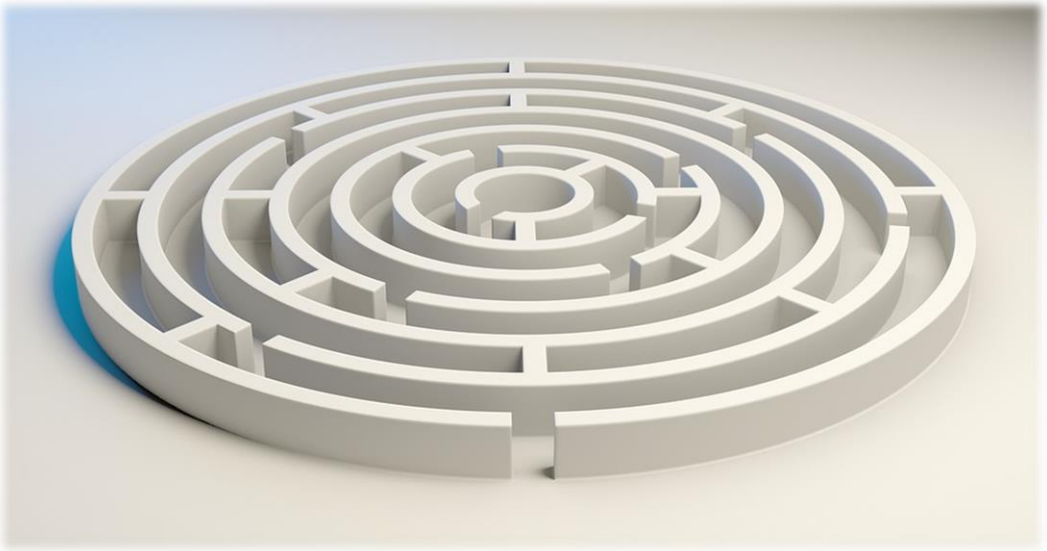
Optimisation

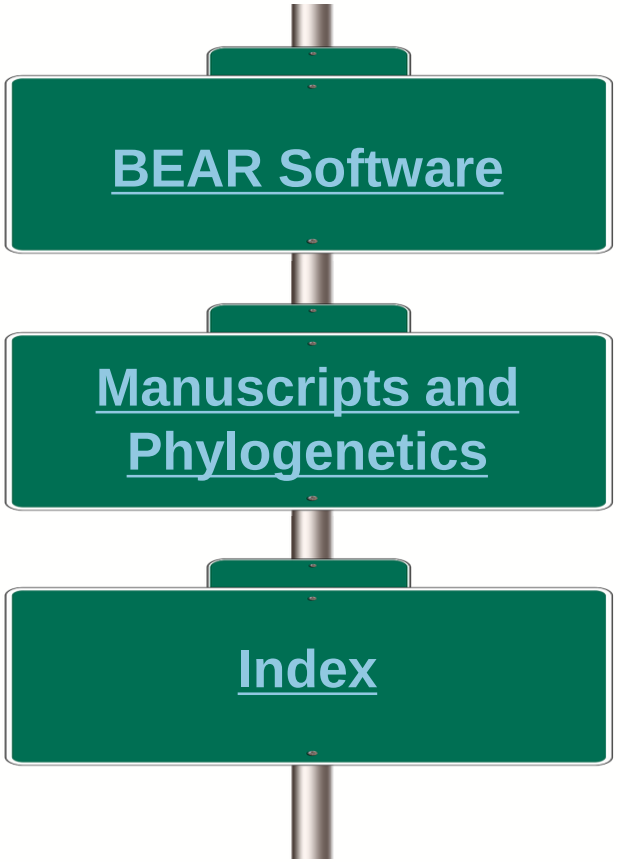
- Profiling
  - Language/system specific
- DRY
- Docstrings
- Architecture documentation
- Then optimise the slow, frequently used bits
  - Optimising usually adds complexity

Complexity

Good Coding Practices

- Questions?




BEAR Software
Manuscripts and Phylogenetics
Index

University of Birmingham

Good Coding Practices

BEAR Software

RSEConUK 2019

IBM POWER9 AI Cluster

Manuscripts and Phylogenetics

EasyBuild

Start again

Pizza and Beer

UNIVERSITY OF BIRMINGHAM | RESEARCH SOFTWARE GROUP