Download considered harmful implies a flexible federation framework

Malcolm Atkinson Malcolm.Atkinson@ed.ac.uk

THE UNIVERSITY of EDINBURGH
informatics




THE UNIVERSITY of EDINBURGH




grnet

Fraunhofer


Transition in data-powered research

- Technical challenges dominated
  *so **we** ignored human issues*

- The power of today's technology
- Plus our immense wealth of data
- Plus the complexity of pressing global challenges
  *means **we** must change **our** behaviour*

Transition in data-powered research

- Technical challenges dominated
  *so **we** ignored human issues*

In the past

- The power of today's technology
- Plus our immense wealth of data
- Plus the complexity of pressing global challenges
  *means **we** must change **our** behaviour*

From now on

Transition in data-powered research

- Technical challenges dominated
  *so **we** ignored human issues*

In the past

- The power of today's technology
- Plus our immense wealth of data
- Plus the complexity of pressing global challenges
  *means **we** must change **our** behaviour*

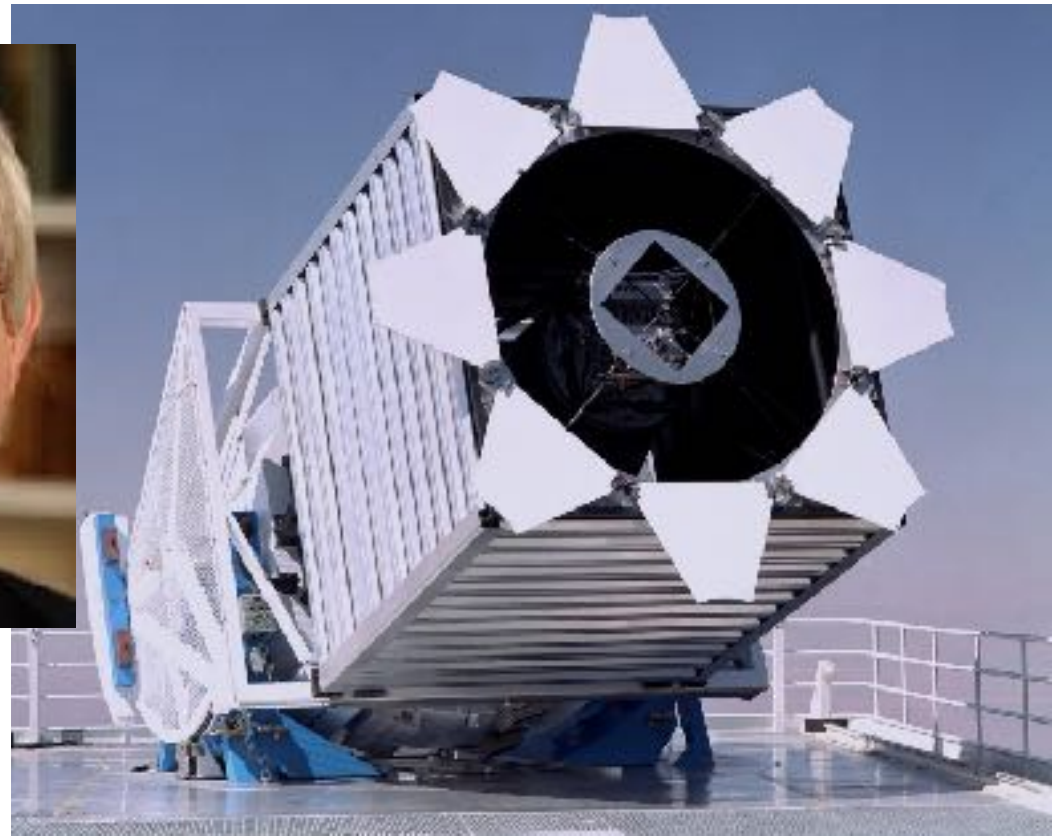
From now on

to empower others to change theirs

Story line

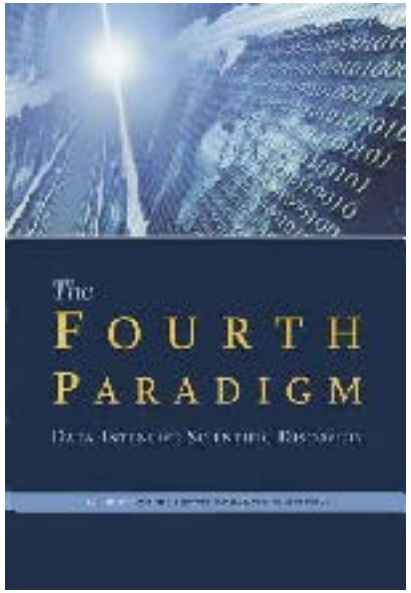
- Emergence of the data bonanza
- Introducing human
  - concerns
  - governance
  - judgement
- Balancing automation with control
  - lingua franca
  - persistent pervasive honouring of intentions
  - respecting all forms of diversity
- Balancing stability with agility

The 4th Paradigm

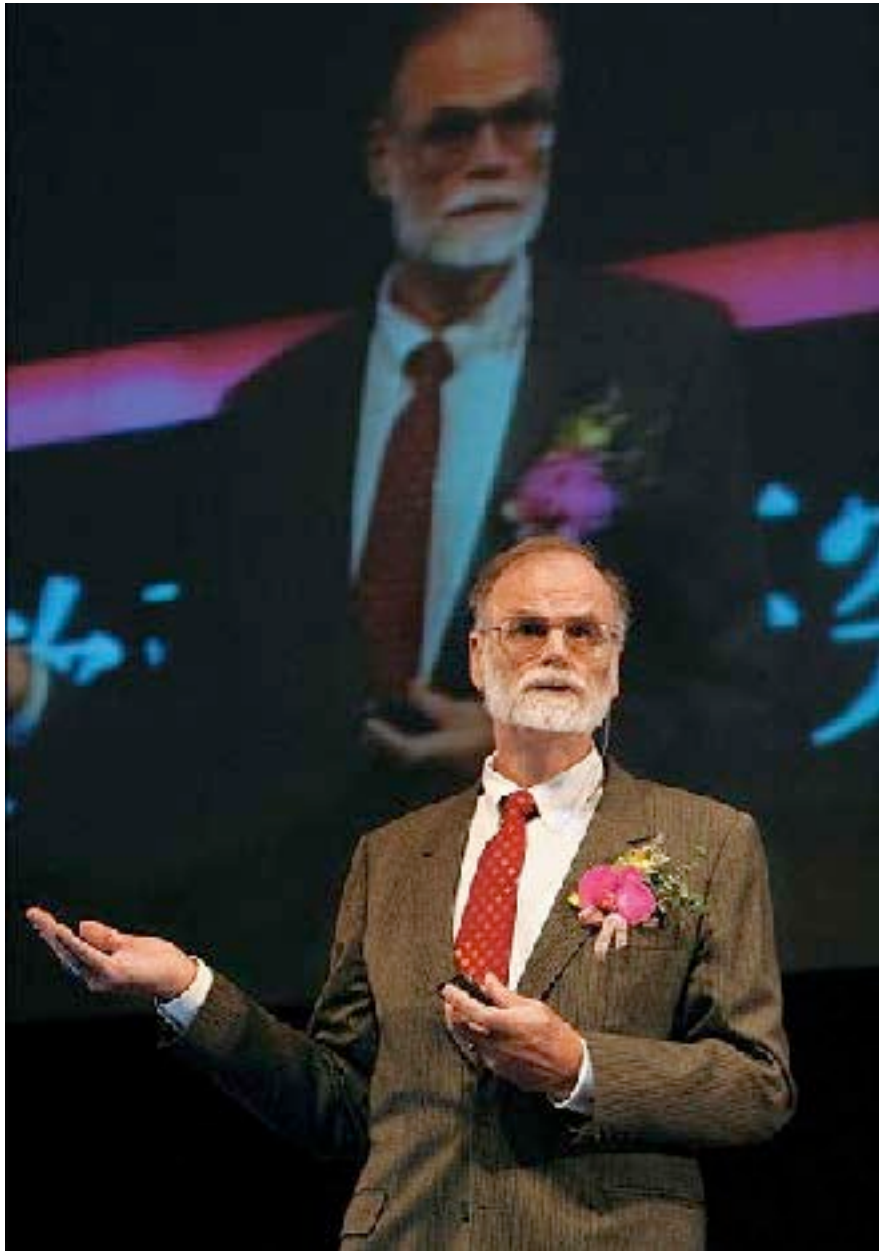


- Experimental Science
- Theoretical Science
- Computational Science
- **Data-intensive Science**
  - Digital technology yields data bonanza
  - Laden with *latent* information
  - Challenges learning
    - to handle volume
    - to discover the knowledge
    - to share the opportunities openly
  - Outruns Moore's law
  - Sociological and technological limits

The 4th Paradigm

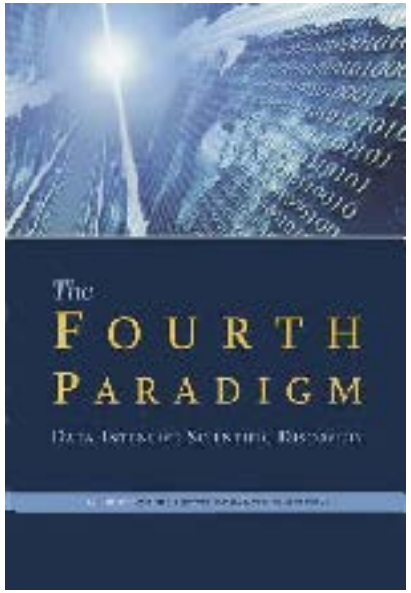


- Experimental Science
- Theoretical Science
- Computational Science
- **Data-intensive Science**
  - Digital technology yields data bonanza
  - Laden with *latent* information
  - Challenges learning
    - to handle volume
    - to discover the knowledge
    - to share the opportunities openly
  - Outruns Moore's law
  - Sociological and technological limits



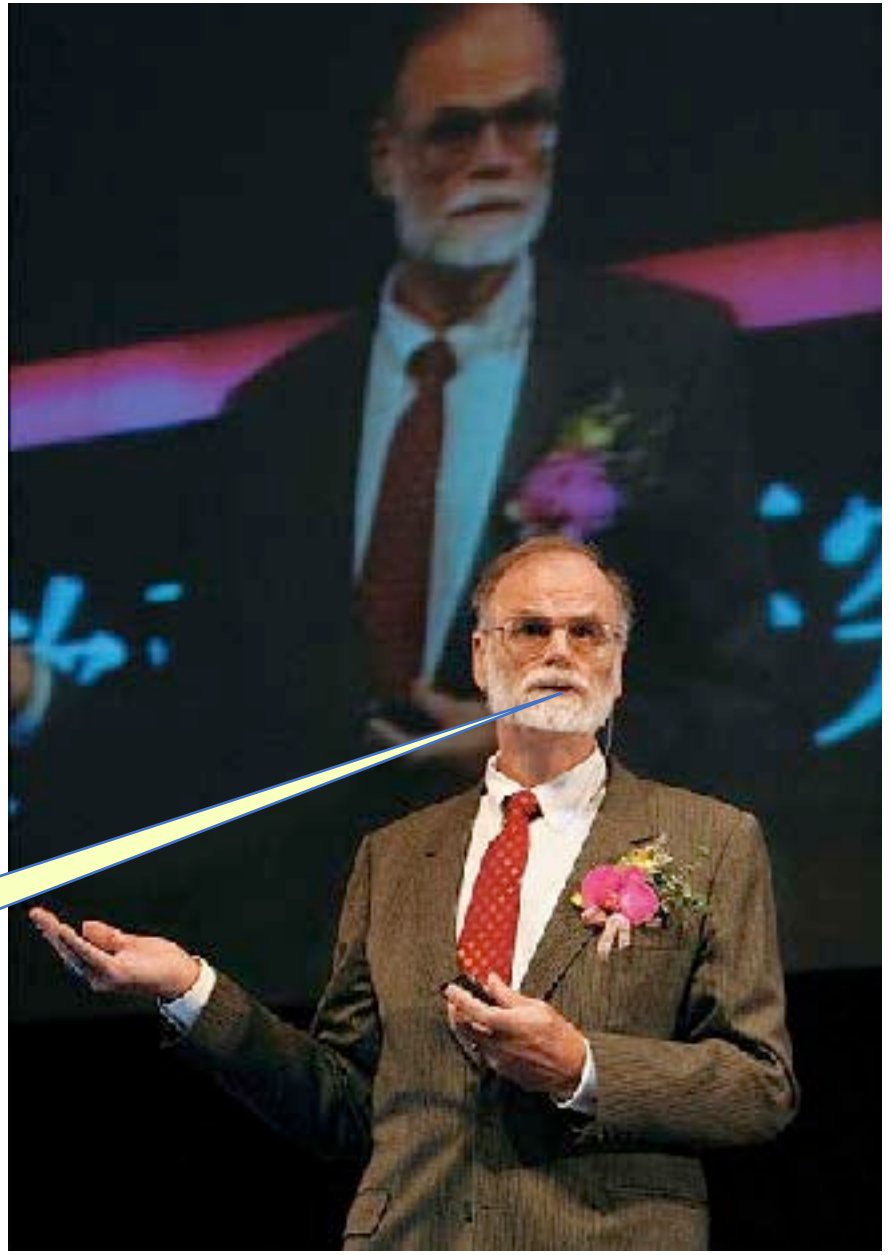
Jim Grey, Microsoft Research

The 4th Paradigm



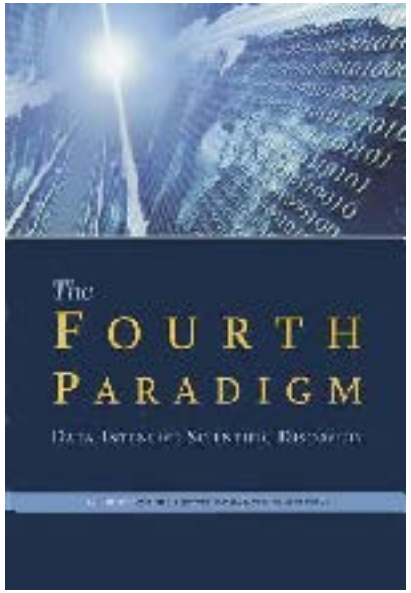
- Experimental Science
- Theoretical Science
- Computational Science
- **Data-intensive Science**
  - Digital technology yields data bonanza
  - Laden with *latent* information
  - Challenges learning
    - to handle volume
    - to discover the knowledge
    - to share the opportunities
  - Outruns Moore's law
  - limits

May all your problems be technical ones



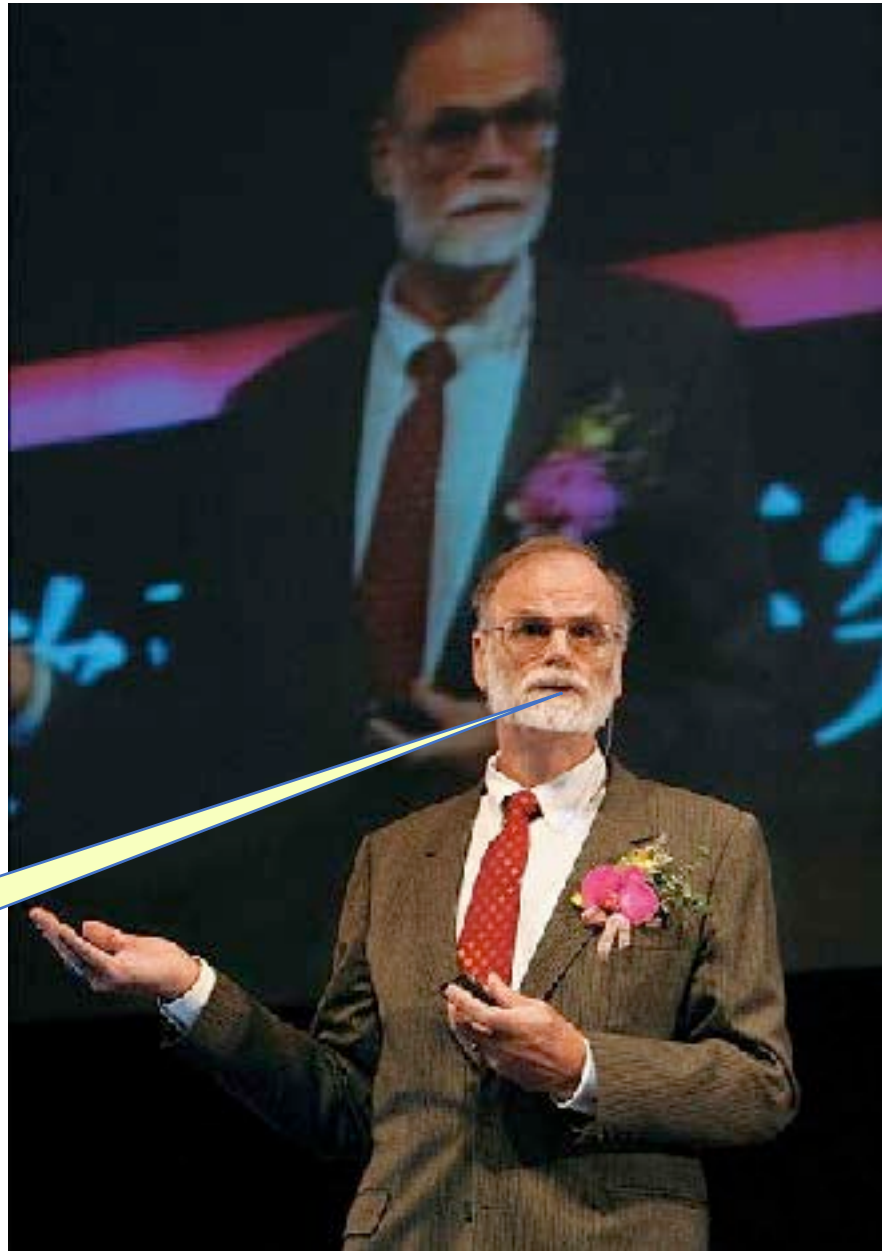
Jim Grey, Microsoft Research

The 4th Paradigm



- Experimental Science
- Theoretical Science
- Computational Science
- **Data-intensive Science**
  - Digital technology yields data bonanza
  - Laden with *latent* information
  - Challenges learning
    - to handle volume
    - to discover the knowledge
    - to share the opportunities
  - Outrun Moore's law
  - limits

May all your problems be technical ones



Jim Grey, Microsoft Research

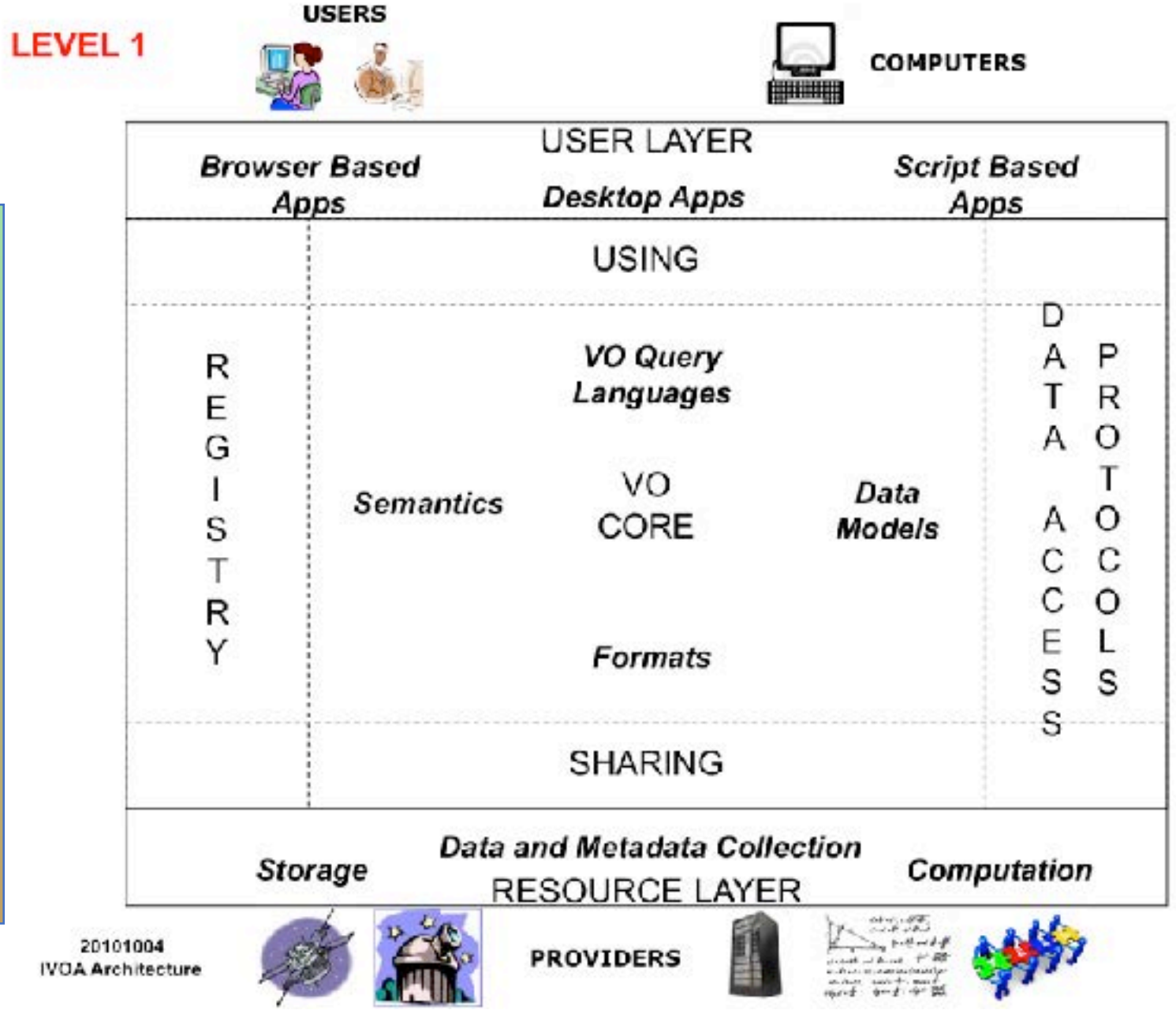
LEVEL 1

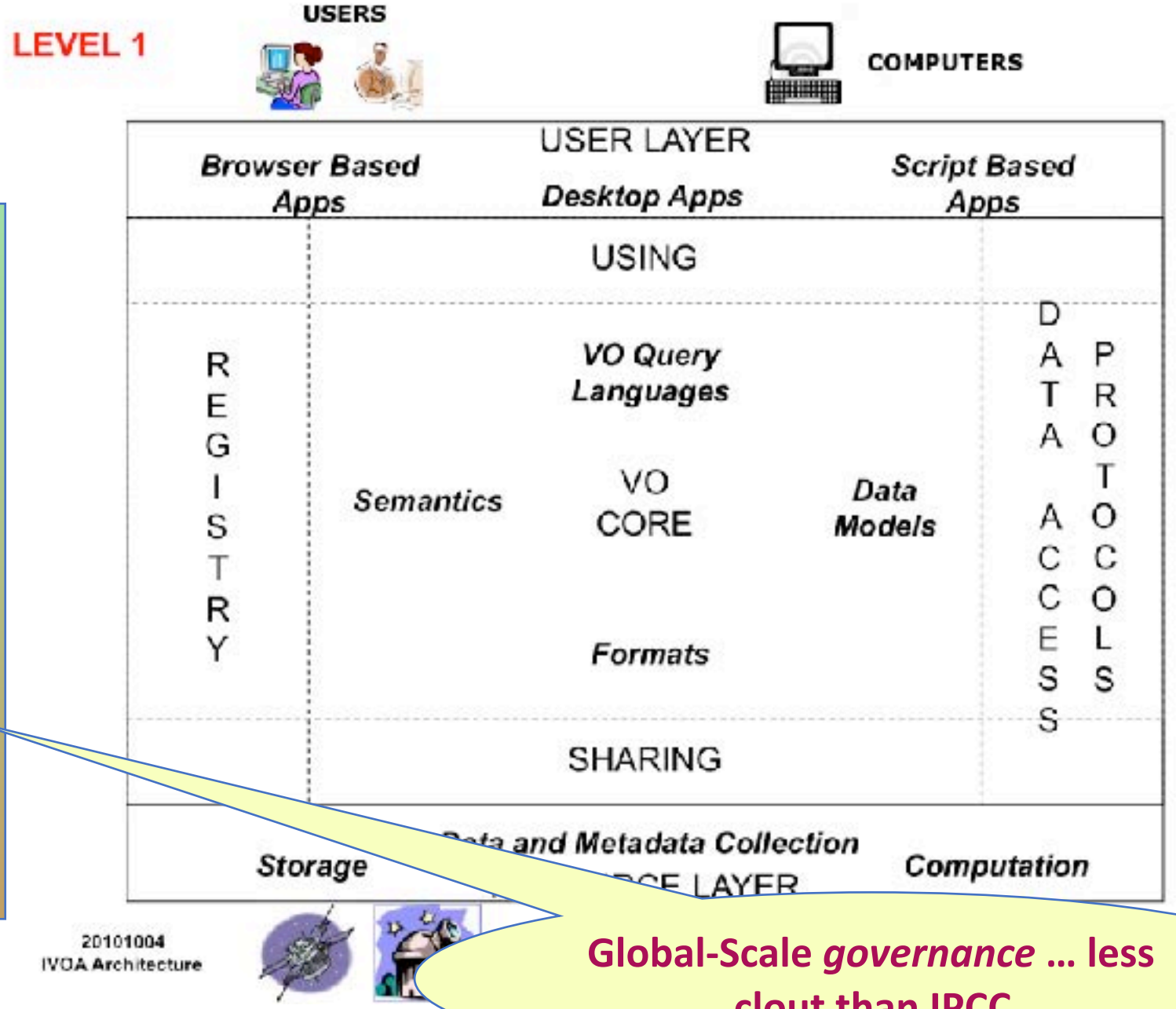
		USER LAYER		
Browser Based Apps		*Desktop Apps*		*Script Based Apps*
		USING		
REGISTRY	*Semantics*	*VO Query Languages*		DATA ACCESS PROTOCOLS
		VO CORE	*Data Models*	
		Formats		
		SHARING		
Storage		*Data and Metadata Collection* RESOURCE LAYER		*Computation*

20101004
IVOA Architecture

PROVIDERS

rapid
development of
agreements
global adoption
continuous
extension



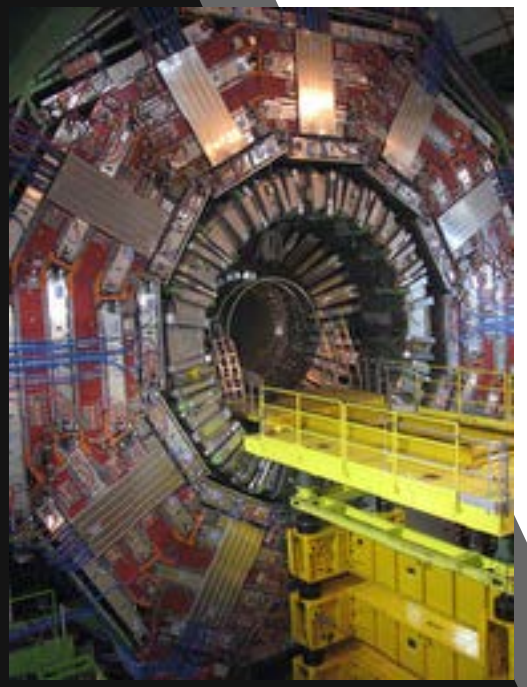


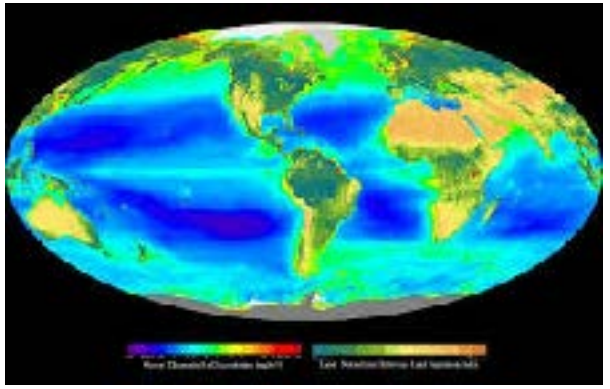
rapid
development of agreements
global **adoption**
continuous extension

Global-Scale *governance* … less clout than IPCC

Examples

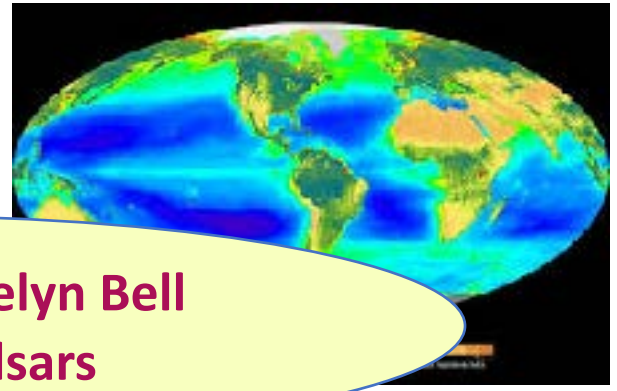
- Astronomy
- Environmental sciences
- Climate and weather
- High-energy physics
- Life sciences
- Social sciences
- Humanities





Examples

- Astronomy
- Environmental sciences
- Climate and weather
- High-energy physics
- Life sciences
- Social sciences
- Humanities



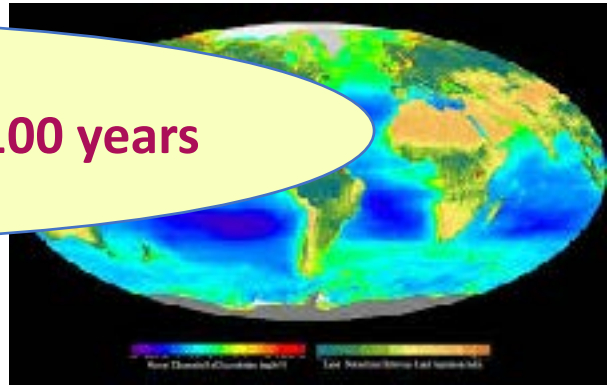
50 years since Jocelyn Bell discovered pulsars

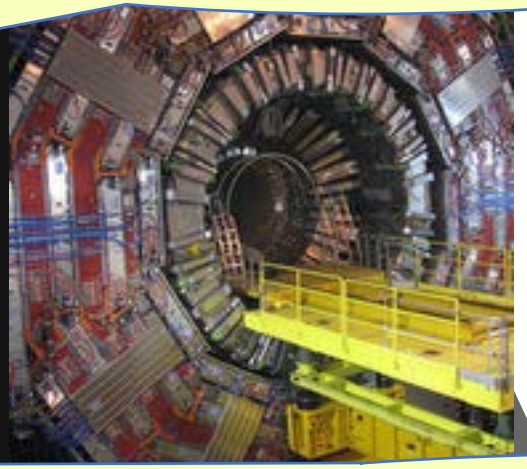
50 years since plate tectonics modelled

Examples

- Astronomy
- Environmental sciences
- Climate and weather
- High-energy physics
- Life sciences
- Social sciences
- Humanities

Gravitational waves after 100 years



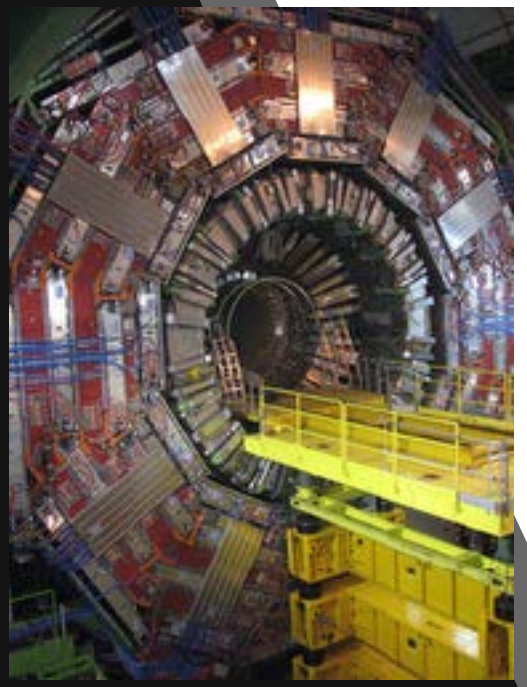


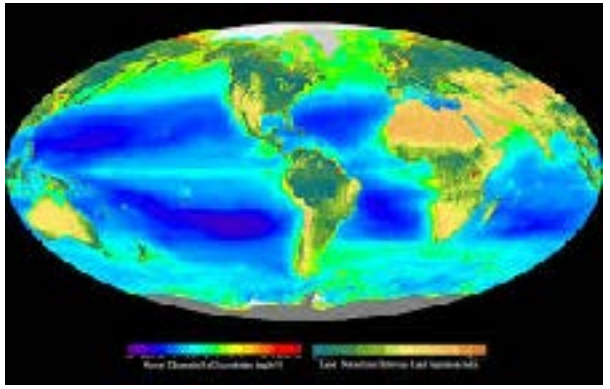
twitter

Higgs boson
after 50 years

Examples

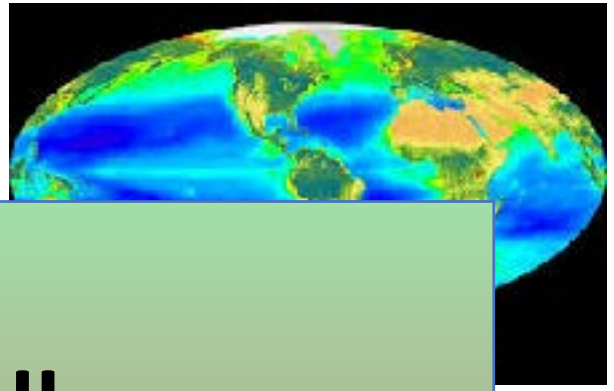
- Astronomy
- **Environmental sciences**
- **Climate and weather**
- High-energy physics
- Life sciences
- Social sciences
- Humanities

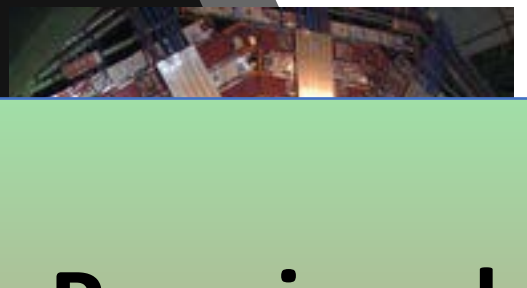




Examples

- Astronomy
- **Environmental sciences**
- **Climate and weather**
- High-energy physics
- Life sciences
- Social sciences
- Humanities





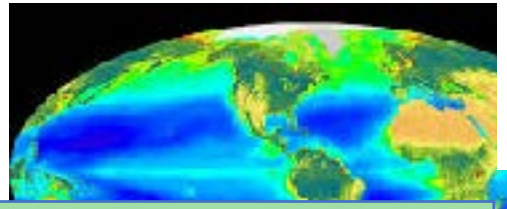
Pressing challenges

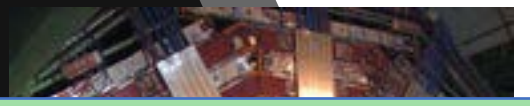
with

global and local pressures

Examples

- Astronomy
- **Environmental sciences**
- **Climate and weather**
- High-energy physics
- Life sciences
- Social sciences
- Humanities





International
Commercial
Forces

Pressing challenges

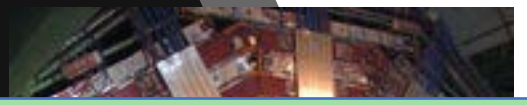
with

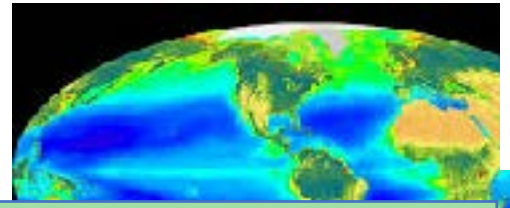
global and local pressures

Example

- Astronomy
- **Environmental sciences**
- **Climate and weather**
- High-energy physics
- Life sciences
- Social sciences
- Humanities

Retaliation
Biases
Tradition
Partition





International
Commercial
Forces

Pressing challenges

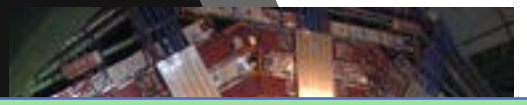
with

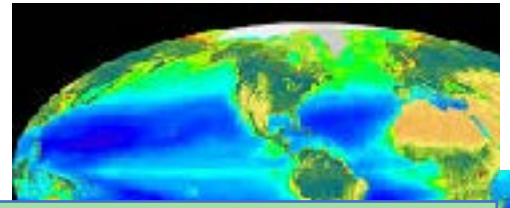
global and local pressures

Example

- Astronomy
- **Environmental sciences**
- **Climate and weather**
- High-energy physics
- Life sciences
- Social sciences
- Humanities

Retaliation
Biases
Tradition
Partition





International
Commercial
Forces

Pressing challenges

with

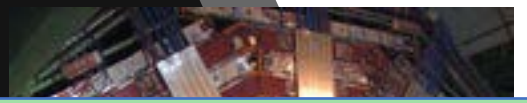
global and local pressures

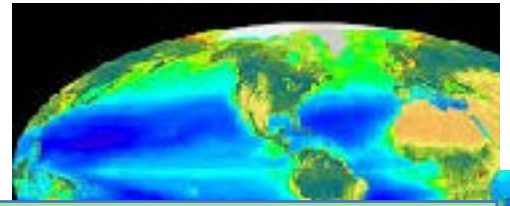
Lives
Economics
Resources
Values

Example

- Astronomy
- **Environmental sciences**
- **Climate and weather**
- High-energy physics
- Life sciences
- Social sciences
- Humanities

Retaliation
Biases
Tradition
Partition





International
Commercial
Forces

Pressing challenges

with

global and local pressures

Lives
Economics
Resources
Values

Habitability
Food
Water
Species

50 years ago

Edsger Dijkstra

Letters to the Editor

Go To Statement Considered Harmful

Key Words and Phrases: go to statement, jump instruction, branch instruction, conditional clause, alternative clause, repetitive clause, program intelligibility, program sequencing
CR Categories: 4.22, 5.23, 5.24

EDITOR:

For a number of years I have been familiar with the observation that the quality of programmers is a decreasing function of the density of **go to** statements in the programs they produce. More recently I discovered why the use of the **go to** statement has such disastrous effects, and I became convinced that the **go to** statement should be abolished from all "higher level" programming languages (i.e. everything except, perhaps, plain machine code). At that time I did not attach too much importance to this discovery; I now submit my considerations for publication because in very recent discussions in which the subject turned up, I have been urged to do so.

My first remark is that, although the programmer's activity ends when he has constructed a correct program, the process taking place under control of his program is the true subject matter of his activity, for it is this process that has to accomplish the desired effect; it is this process that in its dynamic behavior has to satisfy the desired specifications. Yet, once the program has been made, the "making" of the corresponding process is delegated to the machine.

dynamic progress is only characterized when we also g
call of the procedure we refer. With the inclusion of
we can characterize the progress of the process via a
textual indices, the length of this sequence being e
dynamic depth of procedure calling.

Let us now consider repetition clauses (like, **while**
or **repeat** *A* **until** *B*). Logically speaking, such clau
superfluous, because we can express repetition with
recursive procedures. For reasons of realism I don't
clude them: on the one hand, repetition clauses ca
mented quite comfortably with present day finite eq
the other hand, the reasoning pattern known as
makes us well equipped to retain our intellectual g
processes generated by repetition clauses. With the
the repetition clauses textual indices are no longer
describe the dynamic progress of the process. With eac
a repetition clause, however, we can associate a so-
namic index," inexorably counting the ordinal nur
corresponding current repetition. As repetition clau
procedure calls) may be applied nestedly, we find th
progress of the process can always be uniquely charac
(mixed) sequence of textual and/or dynamic indices.

The main point is that the values of these indices
programmer's control; they are generated (either by
of his program or by the dynamic evolution of the proc
he wishes or not. They provide independent coordina

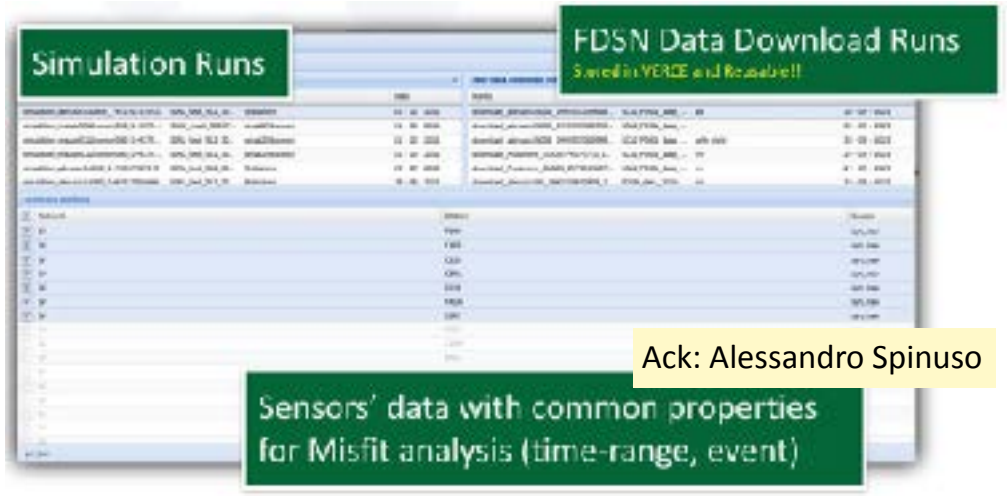
Download considered harmful

Ack: Anusuriya Devaraju

- Prevalent today
- FAIR pushes download
- Evidence of value?
- Researchers own devices
- Provenance unsupported

Download considered harmful

- Prevalent today
- FAIR pushes download
- Evidence of value?
- Researchers own devices
- Provenance unsupported



Ack: Alessandro Spinuso

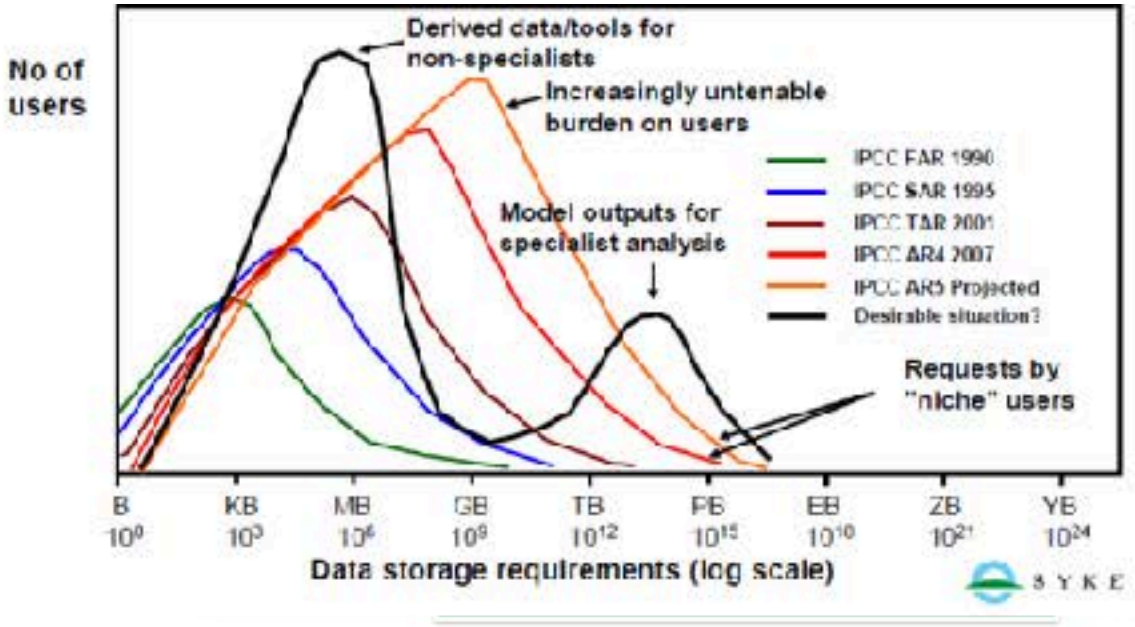
Download considered harmful

- Prevalent today
- FAIR pushes download

Download locally then Analyze:
a workflow that cannot be sustained

- Climate researchers
- Impact researchers

Ack: Christian Pagé



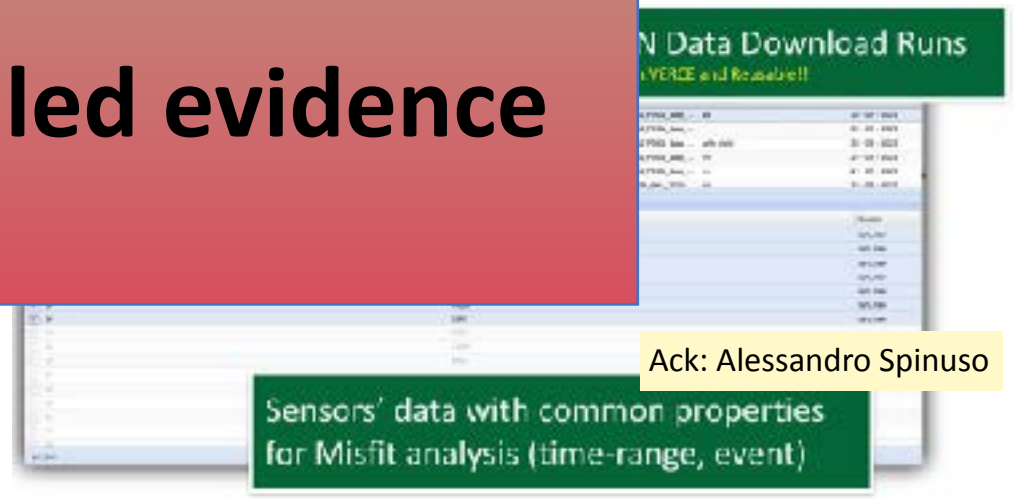
Download considered harmful

- Prevalent today
- FAIR pushes download
- Evidence of value?
- Researchers own devices
- Provenance unsupported

Evidence led policy

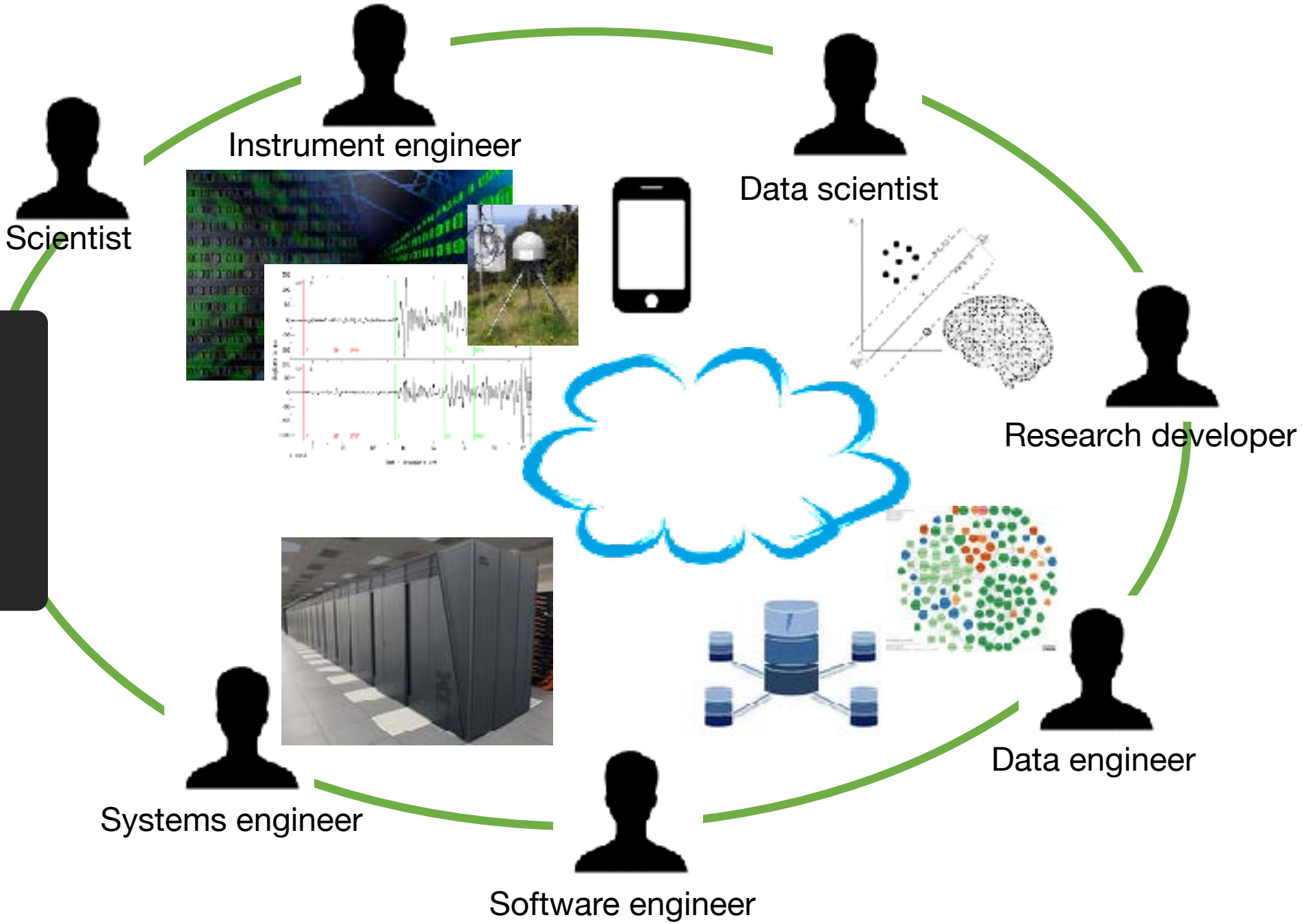
not

policy led evidence

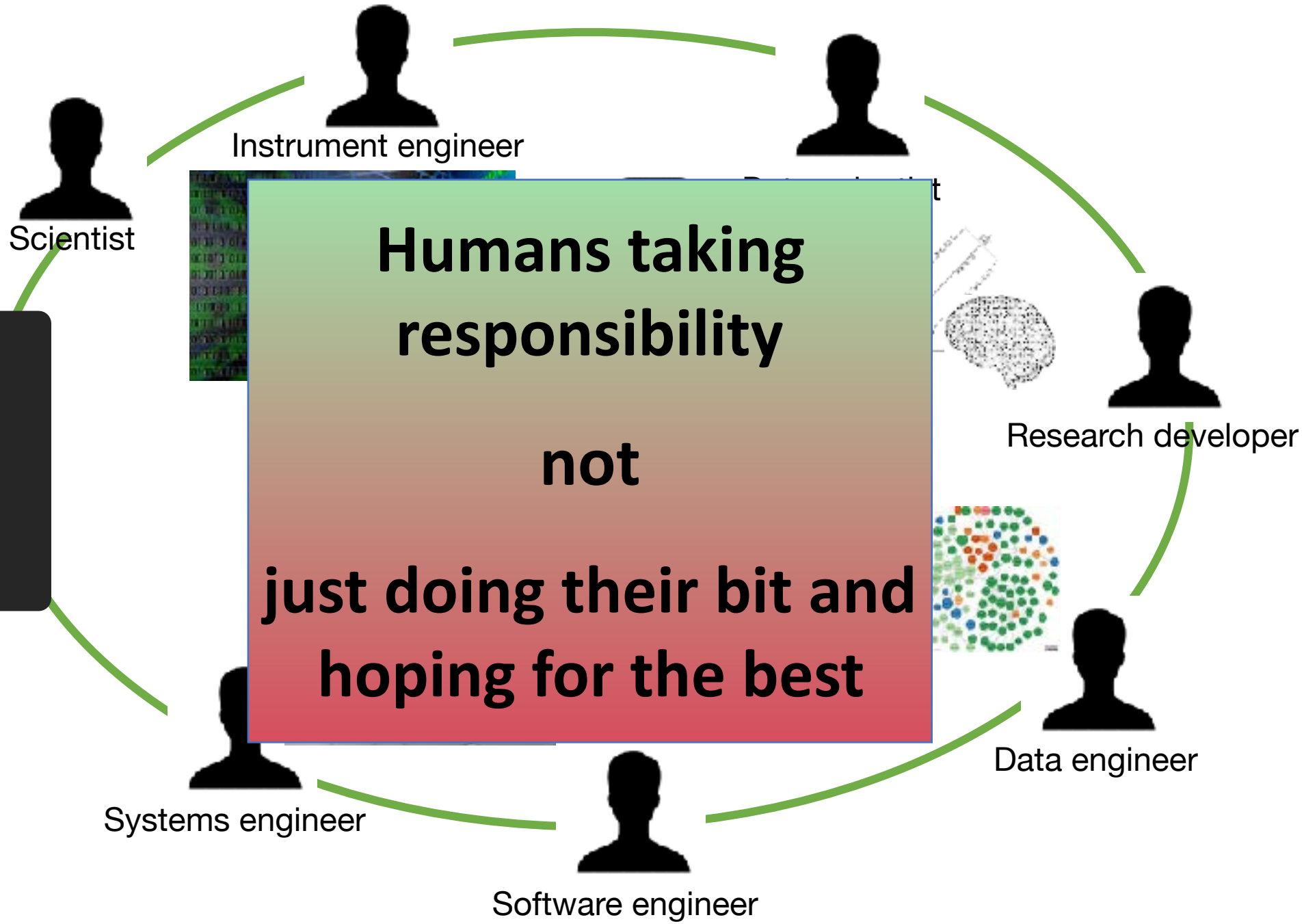


Ack: Alessandro Spinuso

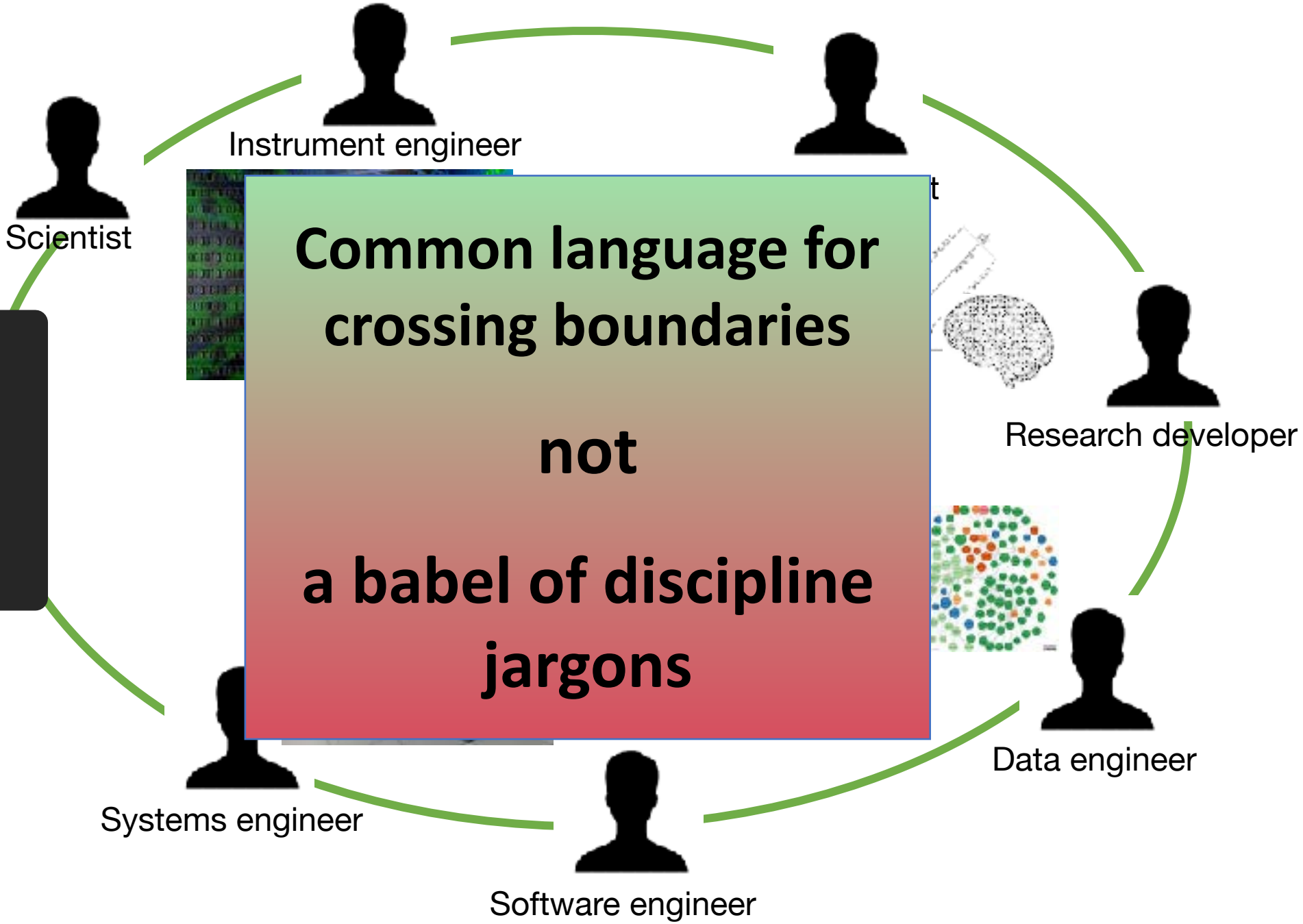
Pooling expertise across disciplines



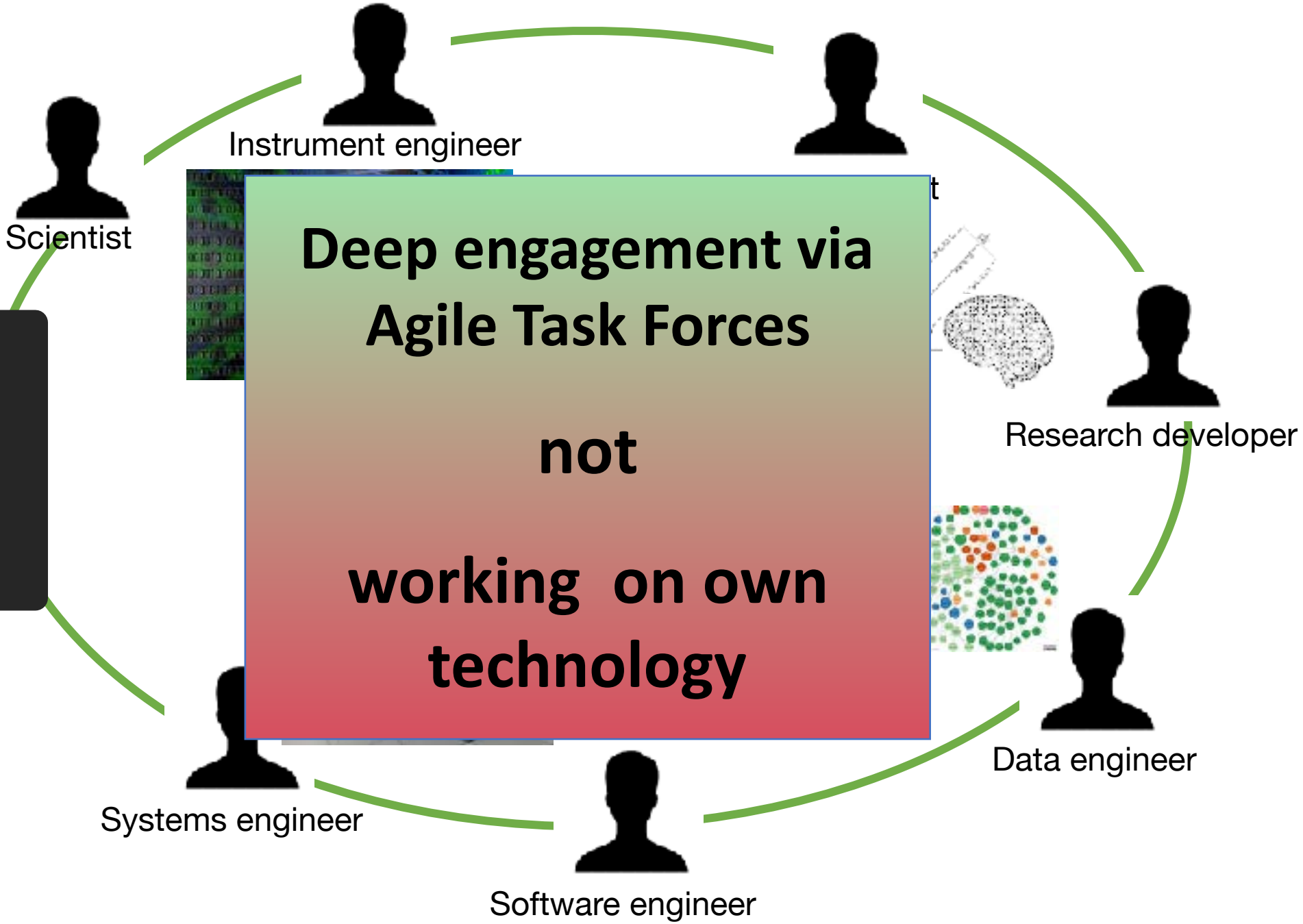
Pooling expertise across disciplines



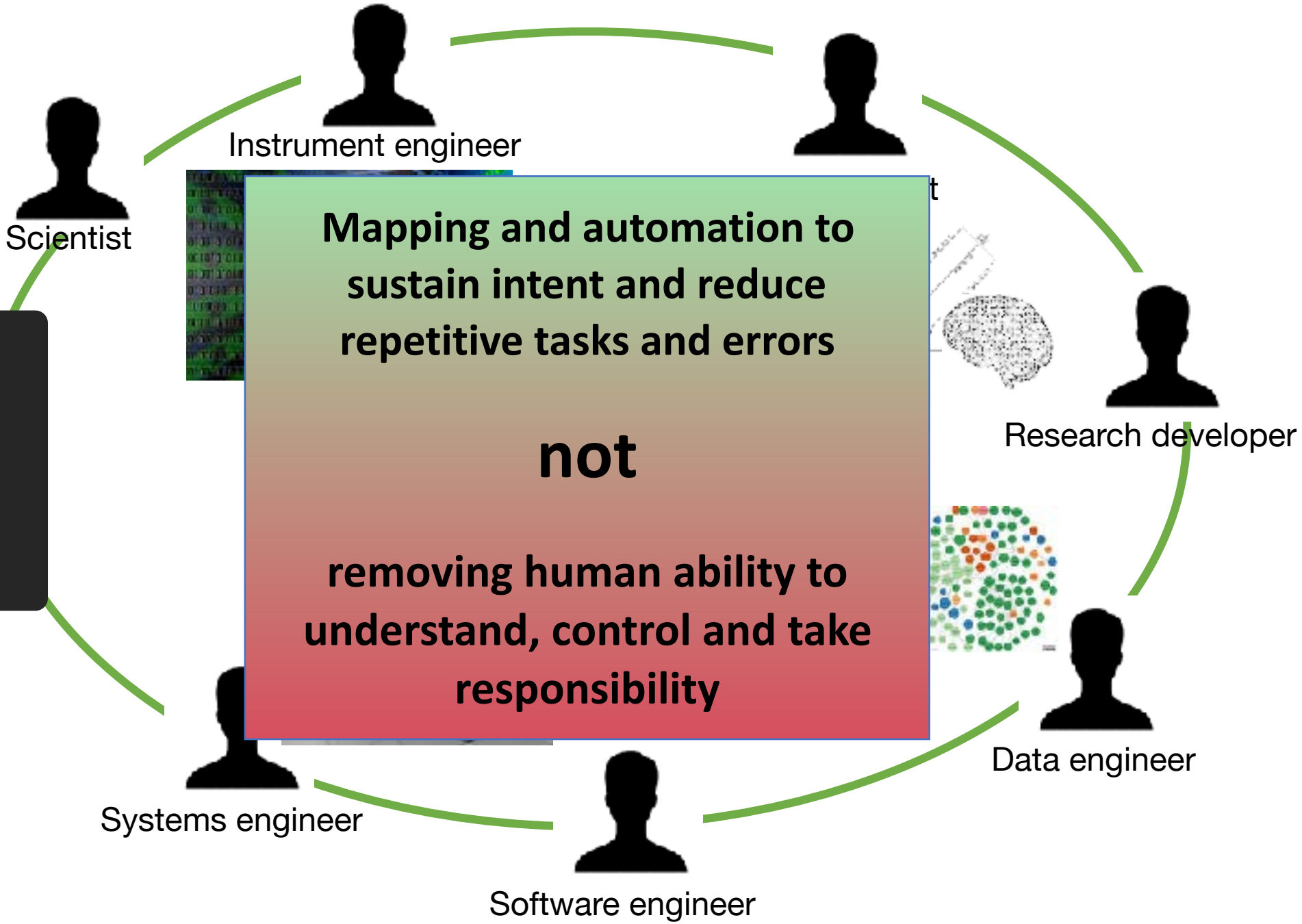
Pooling expertise across disciplines



Pooling expertise across disciplines



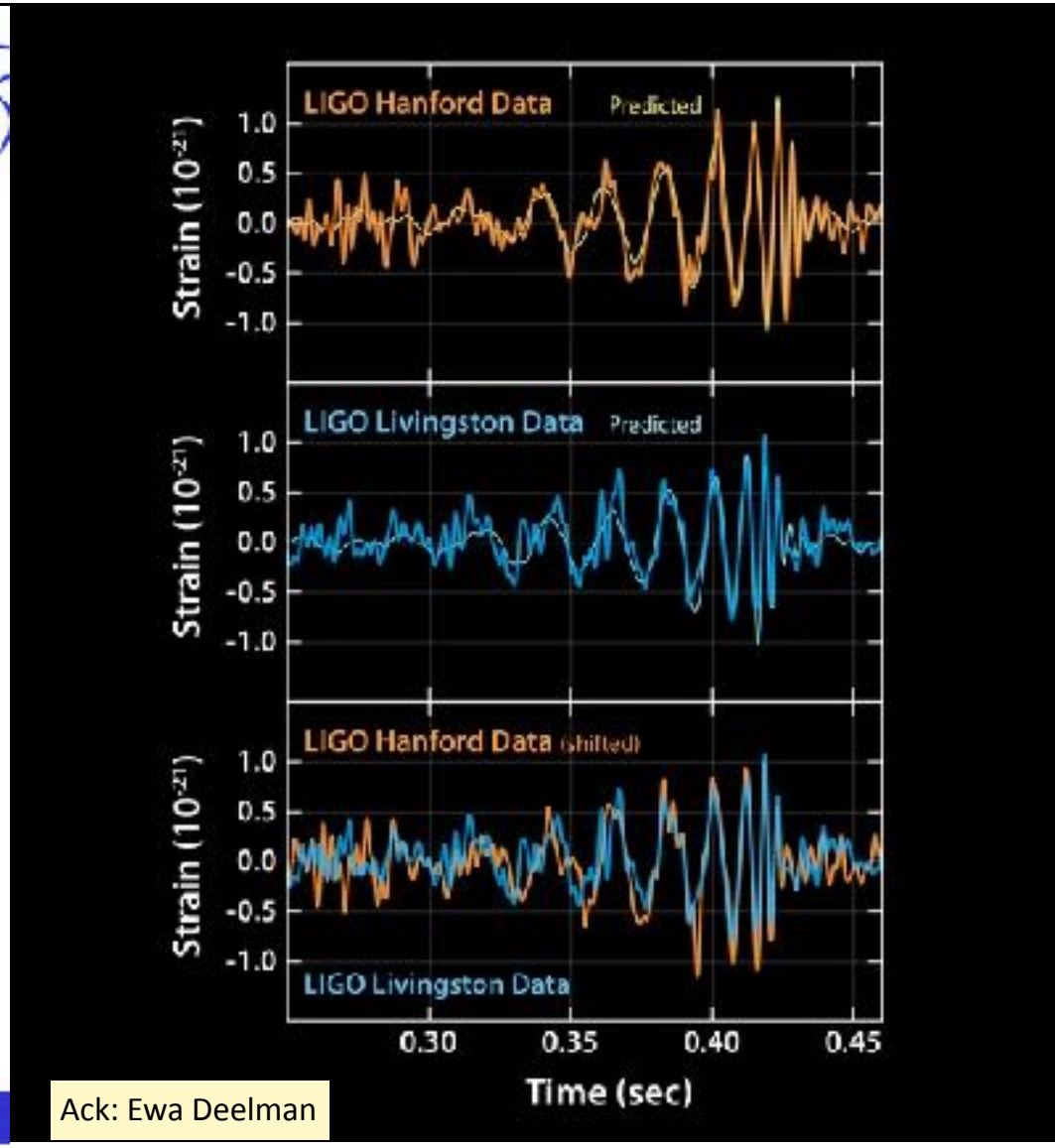
Pooling expertise across disciplines



Gravitational wave detection

LIGO: (Laser Interferometer Gravitational-Wave Observatory)

- Aims to detect gravitational waves predicted by Einstein's theory of relativity.
- Can be used to detect
  - binary pulsars
  - mergers of black holes
  - "starquakes" in neutron stars
- Two installations: in Louisiana (Livingston) and Washington State
  - Other projects: Virgo (Italy), GEO (Germany), Tama (Japan)
- Instruments are designed to measure the effect of gravitational waves on test masses suspended in vacuum.
- Data collected during experiments is a collection of time series (multi-channel)

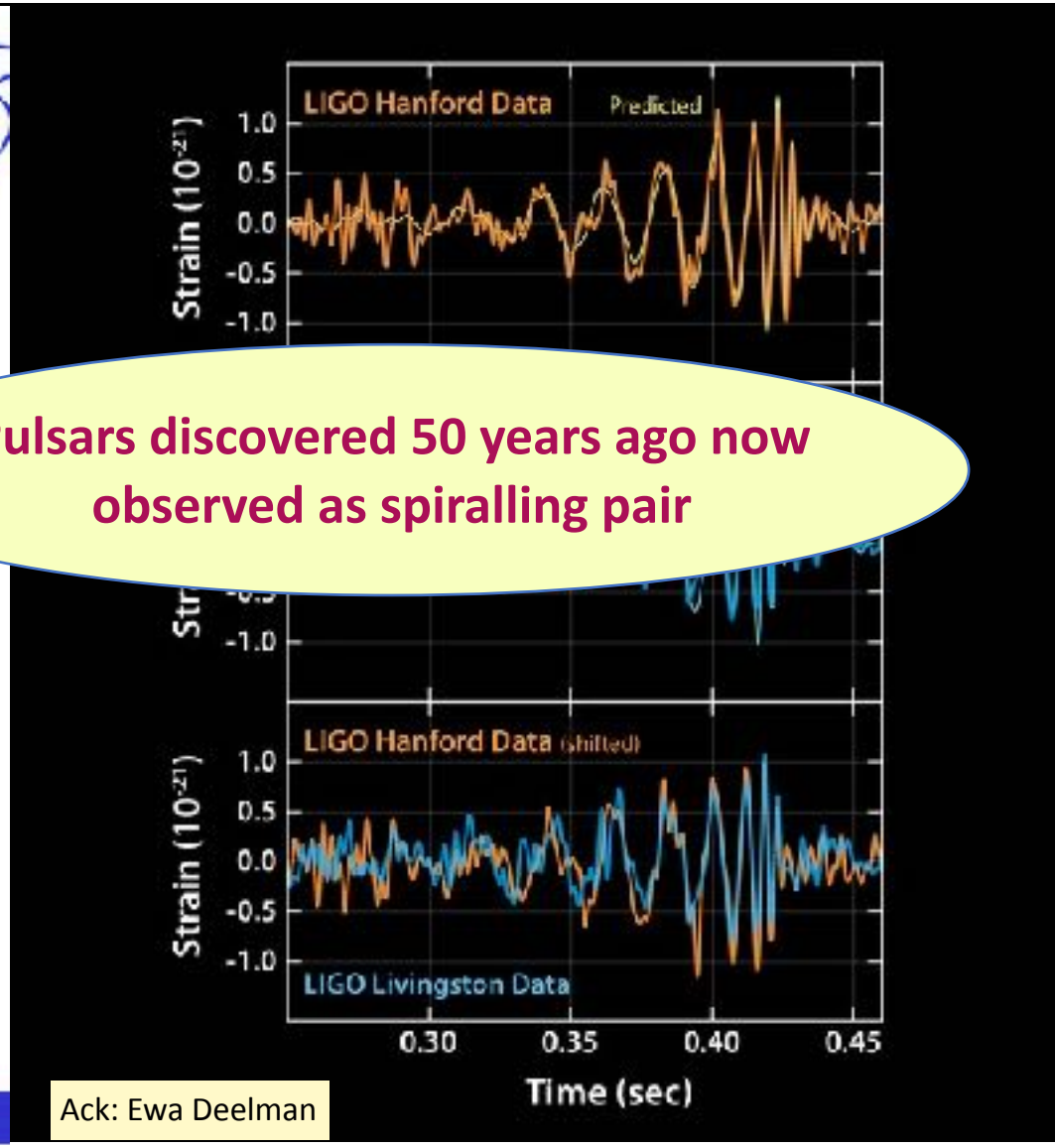


Ack: Ewa Deelman

Gravitational wave detection

LIGO: (Laser Interferometer Gravitational-Wave Observatory)

- Aims to detect gravitational waves predicted by Einstein's theory of relativity.
- Can be used to detect
  - binary pulsars
  - mergers of black holes
  - "starquakes" in neutron stars
- Two installations: in Louisiana (Livingston) and Washington State
  - Other projects: Virgo (Italy), GEO (Germany), Tama (Japan)
- Instruments are designed to measure the effect of gravitational waves on test masses suspended in vacuum.
- Data collected during experiments is a collection of time series (multi-channel)



Pulsars discovered 50 years ago now observed as spiralling pair

Ack: Ewa Deelman

Effective sharing embracing diversity

- Many **autonomous** organisations most engaged in many challenges
- Cluster of **very long** research campaigns
- Scale and scope determined by **challenges**
- Every form of diversity plus **global**
  four Vs for Humans!

Effective sharing embracing diversity

- Many **autonomous** organisations most engaged in many challenges
- Cluster of **very long** research campaigns
- Scale and scope determined by **challenges**
- Every form of diversity plus **global**
  four Vs for Humans!

Reaches scale and breadth for *long-term* clusters of campaigns tackling global and societal challenges

business drives *rapid* change in heterogeneous digital environment

Effective sharing embracing diversity

- Many **autonomous** organisations most engaged in many challenges
- Cluste[illegible]h camp[illegible]
- Scale and scope determined [illegible] **challenges**
- Every form of diversity plus **global** four Vs for Humans!

Darwinian challenge to research campaigns

Reaches scale and breadth for *long-term* clusters of campaigns tackling global and societal challenges

business drives *rapid* change in heterogeneous digital environment

Gaining allegiance incrementally

- Many individuals, roles and priorities
- Established practices plus innovation
- Fair as well as FAIR
- Data, Information, Knowledge, Visualisations, Judgement, Methods, Cultures, Professional practices, QA, Digital platforms, Observational systems, Archives, Computational models, Networks, VREs, Stores, Scientific databases, Ethical rules, Collaboration agreements, DMPs, FAIR, …

Gaining allegiance incrementally

- Many individuals, roles and priorities
- Established practices plus innovation
- Fair as well as FAIR
- Data, Information, Knowledge, Visualisations, Judgement, Methods, Cultures, Professional practices, QA, Digital platforms, Observational systems, Archives, Computational models, Networks, VREs, Stores, Scientific databases, Ethical rules, Collaboration agreements, DMPs, FAIR, …

Building collaboration

without

losing human talent

Gaining allegiance incrementally

- Many individuals, ro
- Established practices
- Fair as well as FAIR
- Data, Information, Knowledge, Visualisations, Judgement, Methods, Cultures, Professional practices, QA, Digital platforms, Observational systems, Archives, Computational models, Networks, VREs, Stores, Scientific databases, Ethical rules, Collaboration agreements, DMPs, FAIR, …

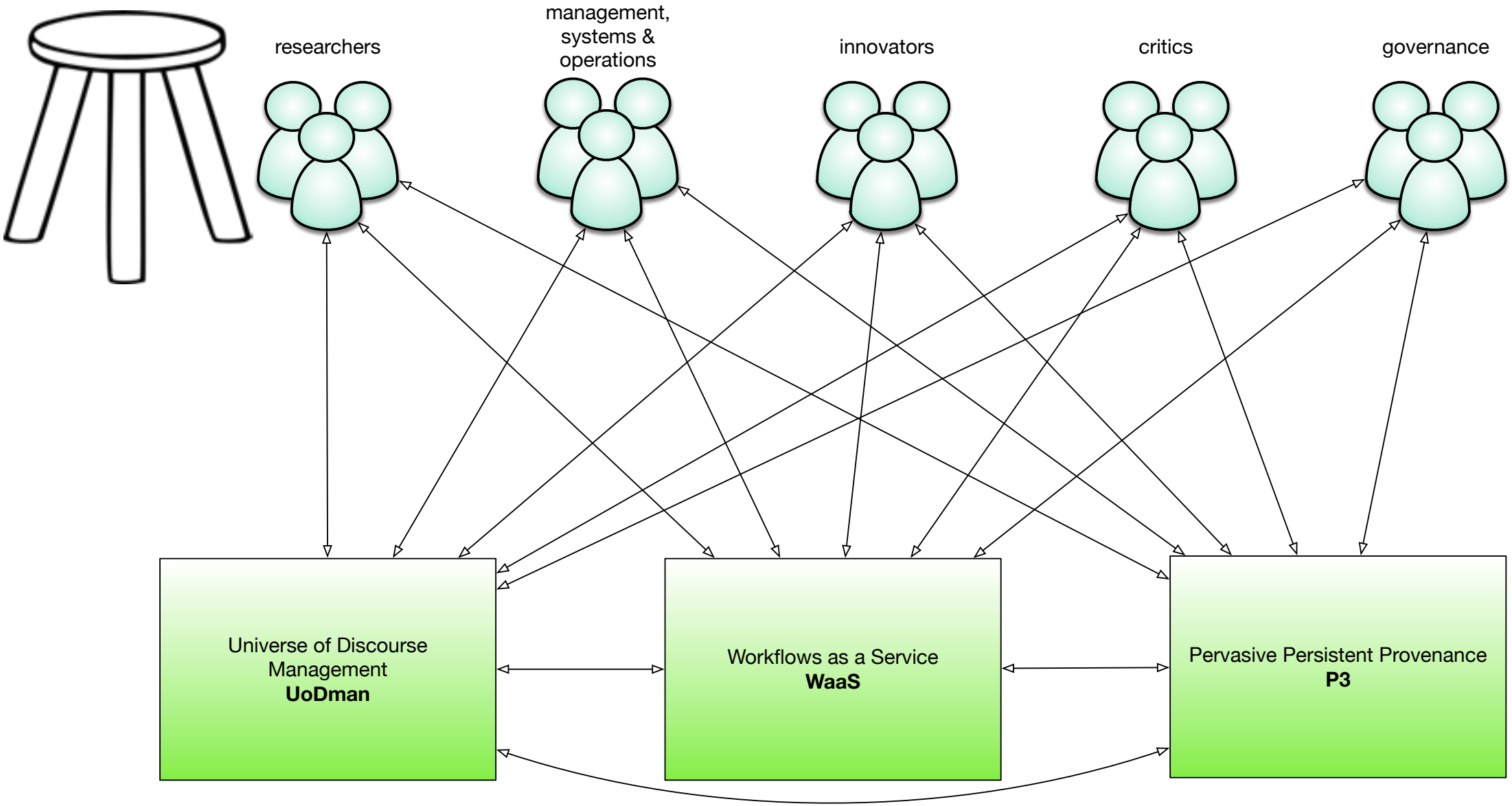
Complexity harnessed via **Integrated understandable controllable abstractions**

Building collaboration

without

losing human talent

Tripos: Flexible Federation Framework (F3)



Roles: one dimension of human variety

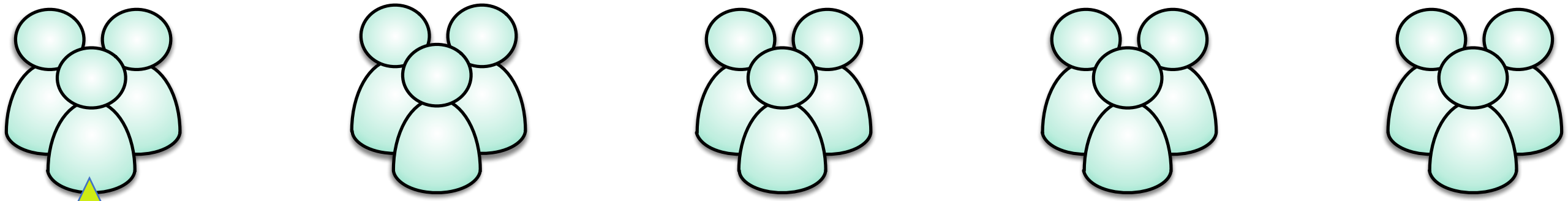
researchers

management, systems & operations

innovators

critics

governance



- 95% professionals 95% time + students + citizen scientists
- Want to keep running their well-developed *research* methods with *their* improvements
- Would like to do more with less effort and faster if possible
- Would like more certainty about the reliability and value of their work
- Worried that imposed changes or collaborators will disrupt their campaigns
- Willing to be helpful and learn new things as long as it doesn't disrupt their work
- 1000s in research campaigns; 100s of specialisms & 10s of disciplines

Roles: one dimension of human variety

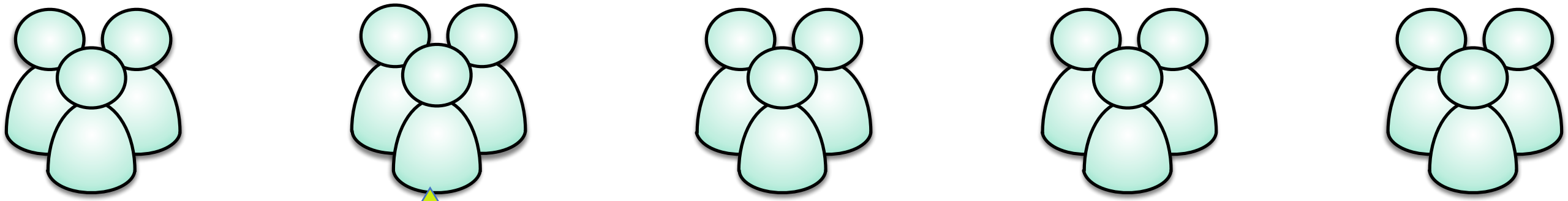
researchers

management, systems & operations

innovators

critics

governance



- Small numbers of mainly experienced professionals some with specialist skills
- Want to keep running their well-developed *operational* methods with *their* improvements
- Would like to do more with less effort and faster and cheaper if possible
- Would like more certainty about the reliability and value of their work
- Worried that things may go wrong while they are in charge
- Have to handle emergencies and maintain long-term stability and gain from trends
- Support all roles in the community and report to stakeholders and governance

Roles: one dimension of human variety

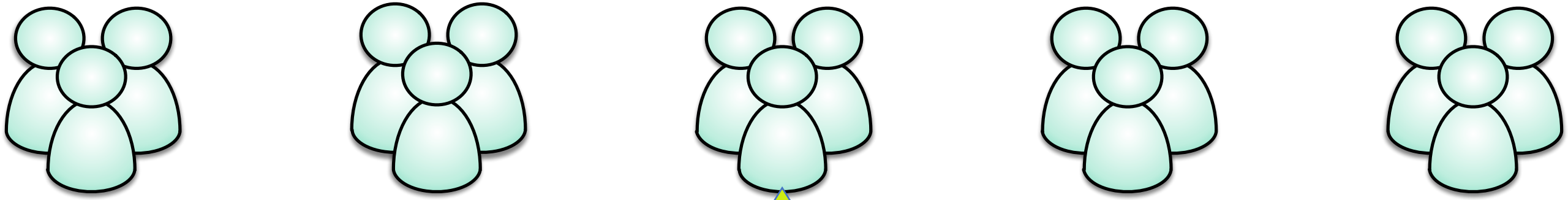
researchers

management, systems & operations

innovators

critics

governance



- 5% of community; research leaders to students; specialists in particular aspects
- Want to solve major challenges and advanced the state of the art
- Want to deliver new capabilities and exploit new opportunities
- Want to repair known weaknesses in the infrastructure, practices or research
- Want to explore freely and not suffer if they don't succeed
- Some have to contribute in emergencies and sustain a long-term strategy
- Expect review of their successes leading to credit, blame and impact

Roles: one dimension of human variety

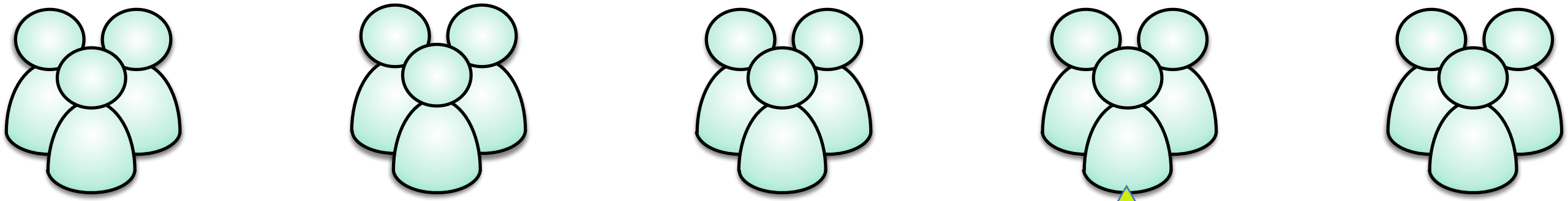
researchers

management, systems & operations

innovators

critics

governance



- External and internal colleagues, with any role and viewpoint: collaborator, rival, sceptic
- Valid review of scientific quality and method
- Threat to reputation, career and community's sustainability
- May expose weaknesses in the organisation, infrastructure, practices or approach
- and weaknesses in any individual entity or collection of entities
- While science should be open there are contexts that need protection from unfair criticism
- Long-term and in-depth study will need resources and (partially automated) support

Roles: one dimension of human variety

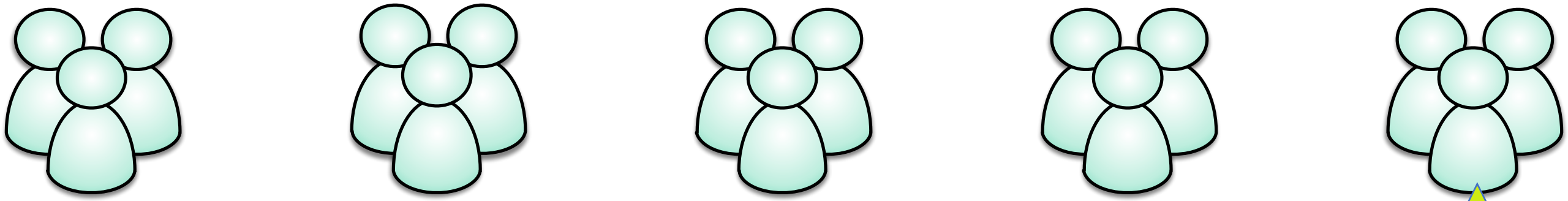
researchers

management, systems & operations

innovators

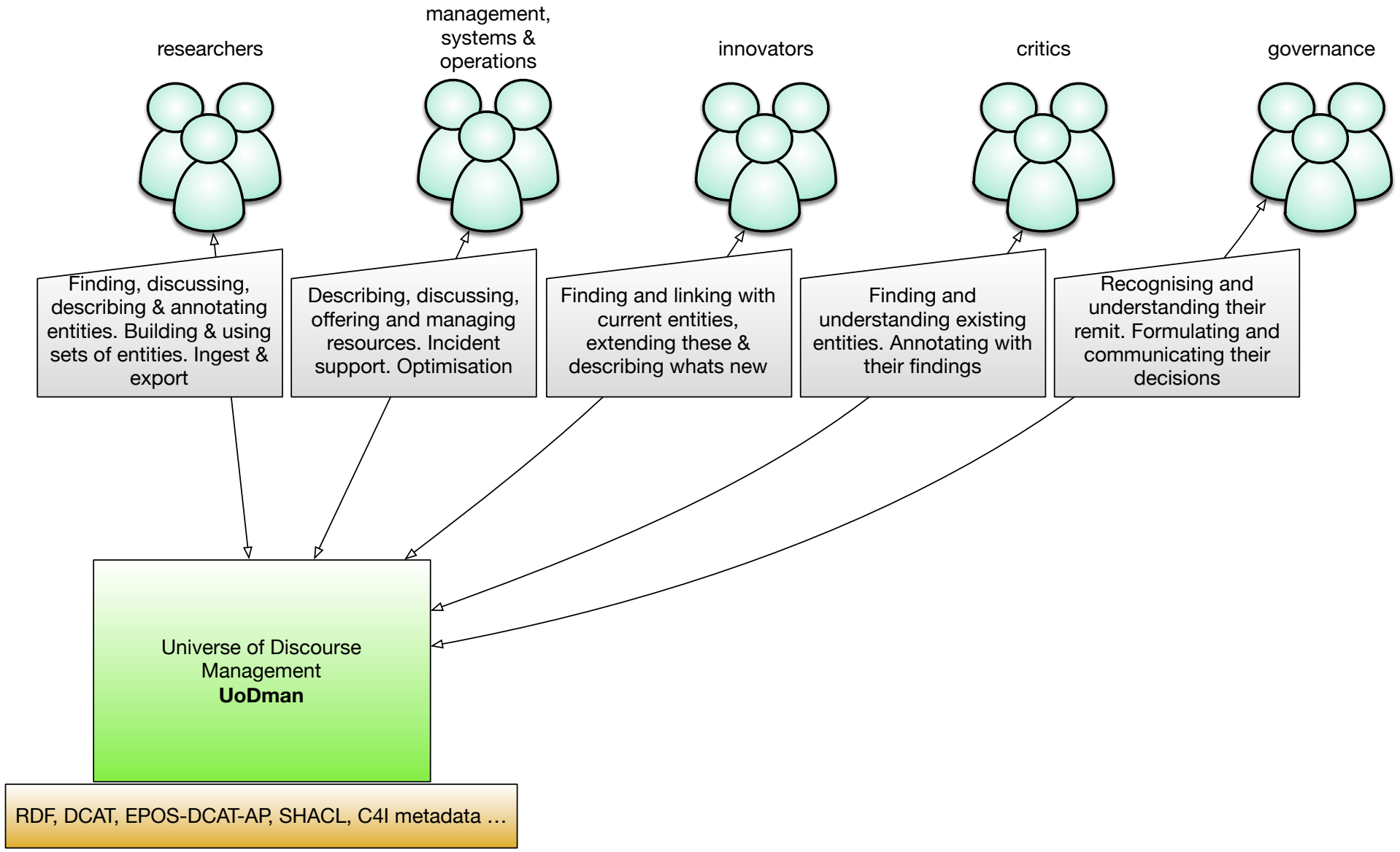
critics

governance

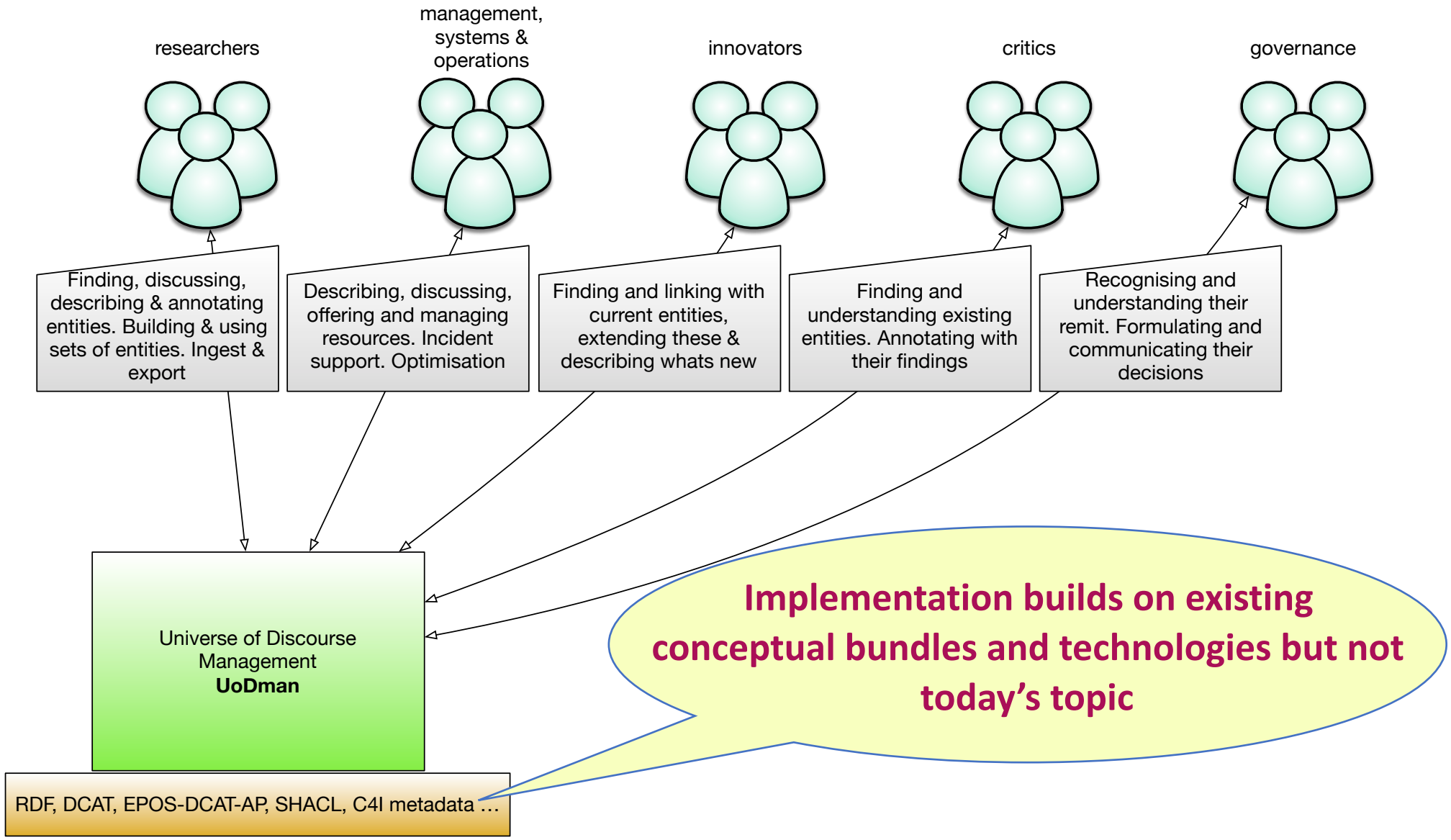


- One or more bodies drawn up from representative stakeholders with access to experts
- Establishing the moral, ethical, and academic ethos and agreed ways of achieving it
- Selecting who to trust and what to trust them with; assigning and monitoring responsibilities
- Meeting the legal and regulatory requirements in each jurisdiction where the federation operates
- Recognising the policies of funders, users and their stakeholders and participant organisations
- Negotiating agreements where these clash
- Propagating the resulting ‘rules’ with clarity to all humans and systems with automated support
- Conducting investigations for appeals and to assess their success

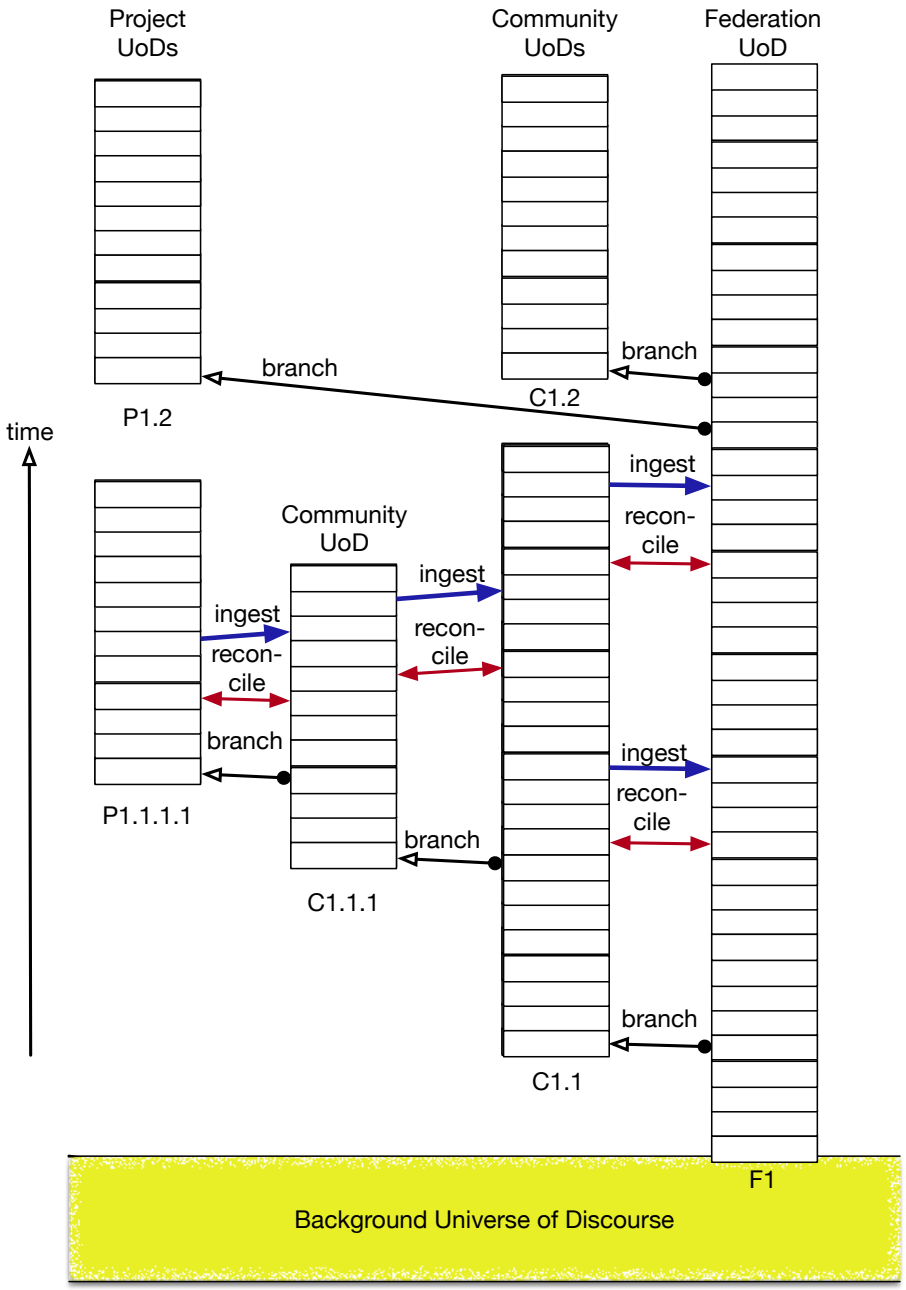
Lingua Franca spanning roles: human & machine



Lingua Franca spanning roles: human & machine

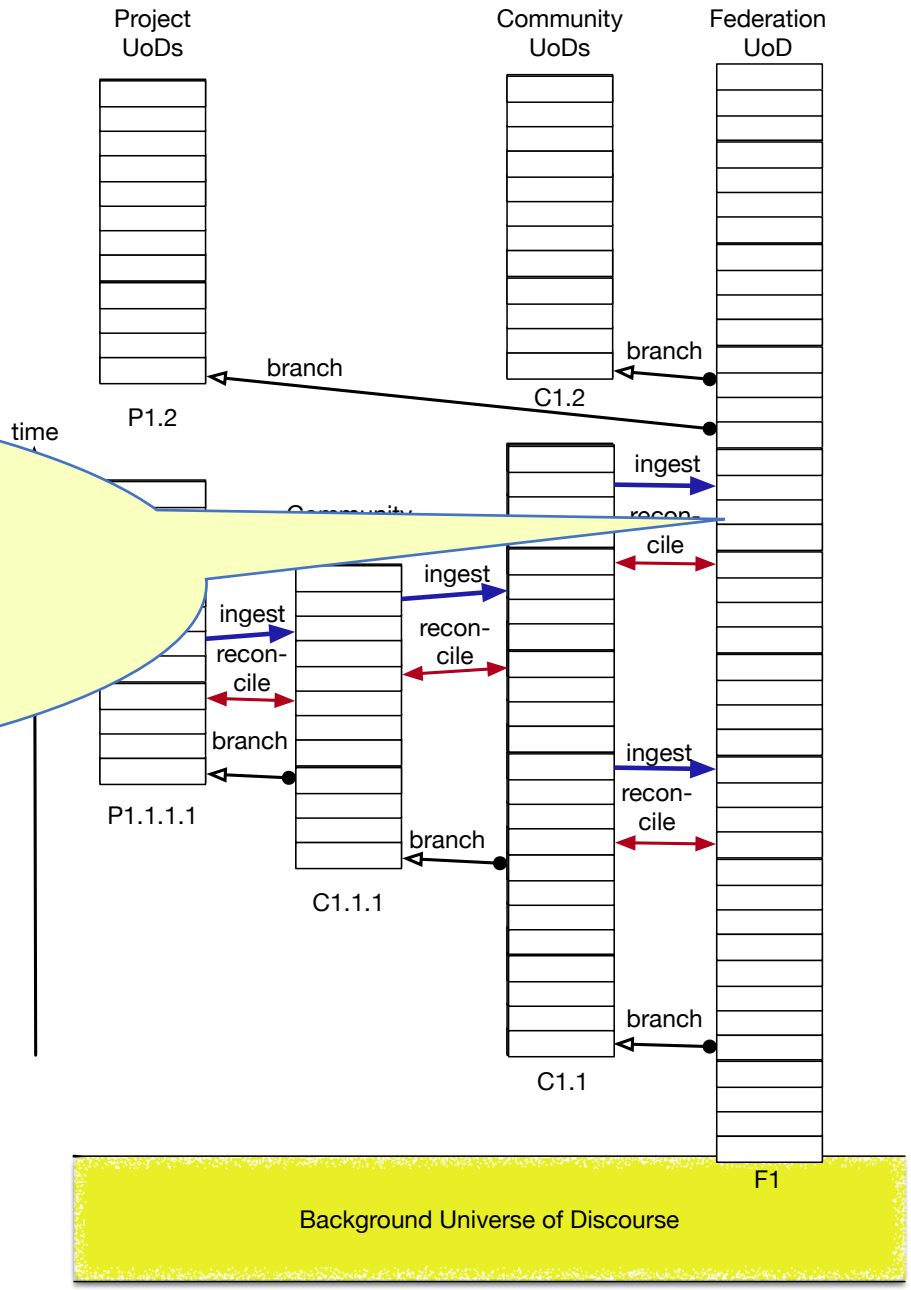


Universes of Discourse



Universes of Discourse

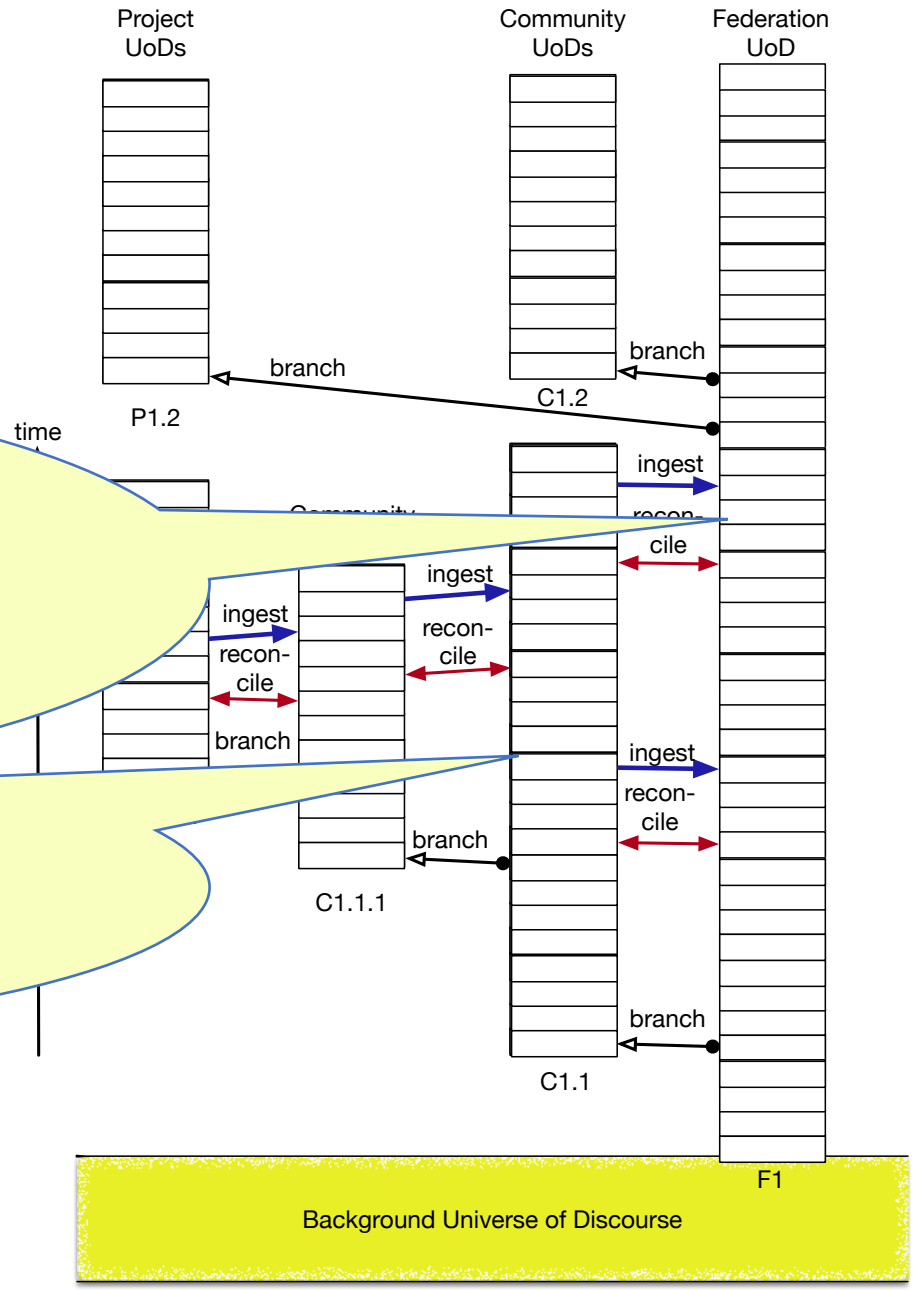
Common *stabilised* concepts, procedures and resources
for
***entire* federation *relative* to background**



Universes of Discourse

Common *stabilised* concepts, procedures and resources
for
***entire* federation *relative* to background**

***UoD* for a significant sustainable sub-community *relative* to common core**

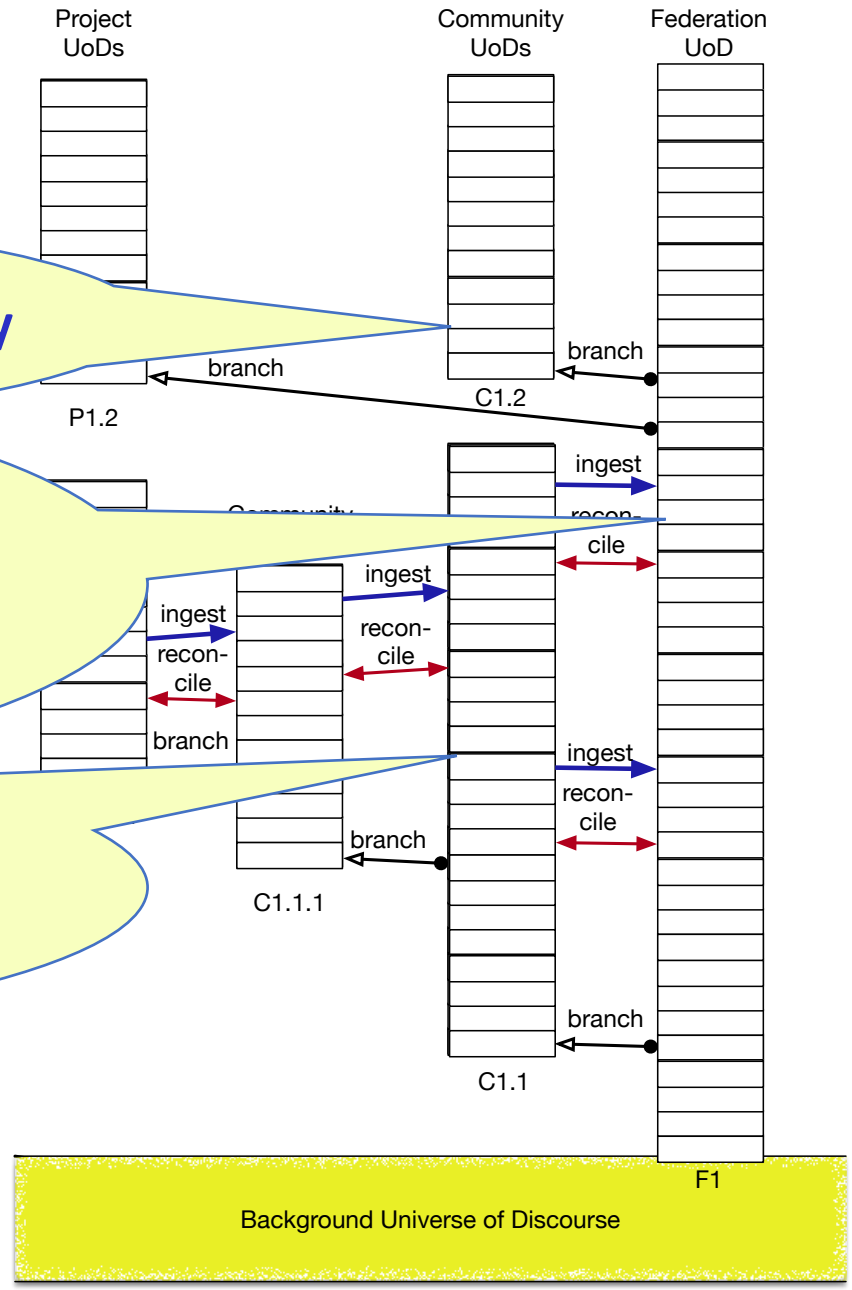


Universes of Discourse

***UoD* for another significant sustainable sub-community**

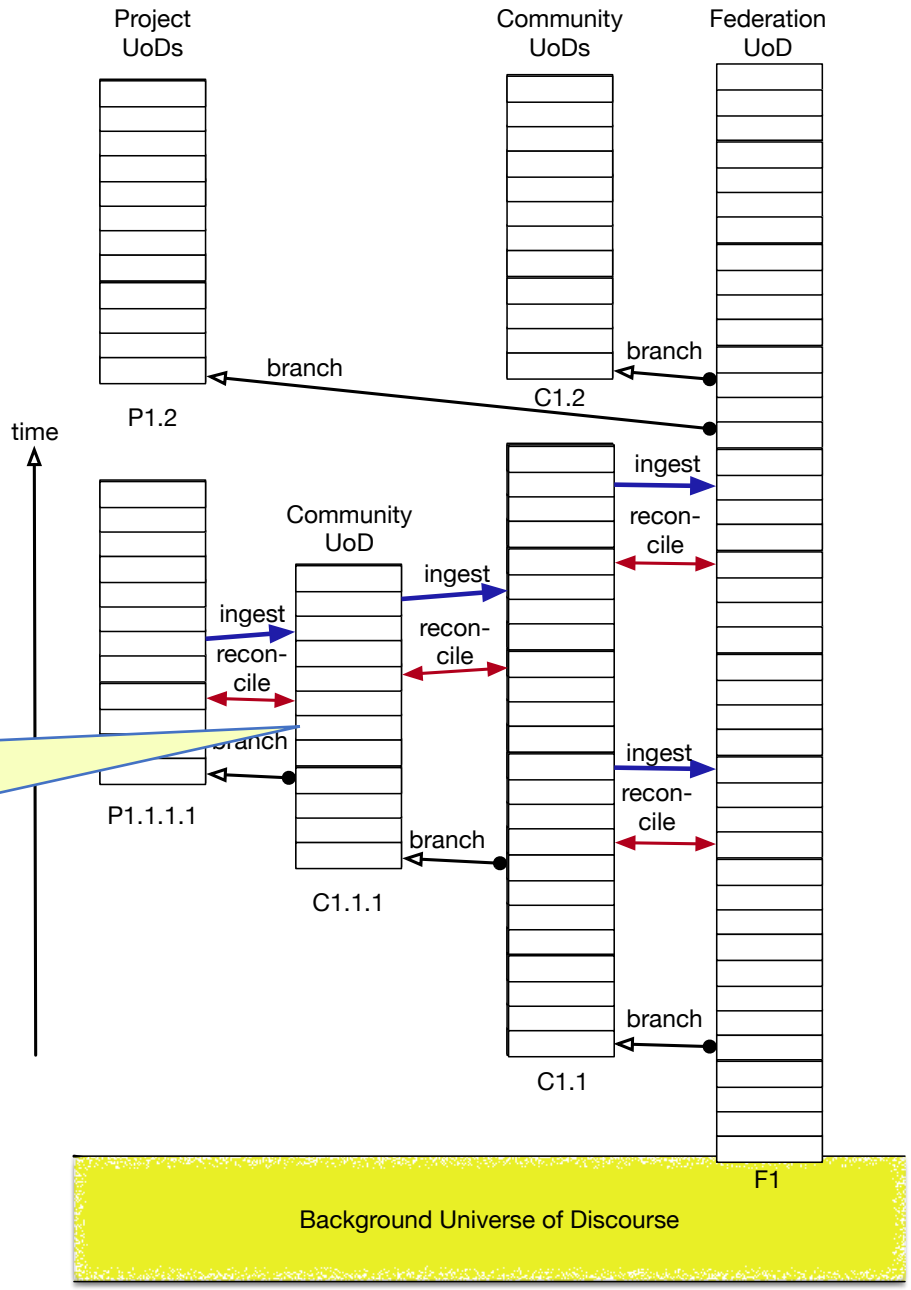
Common *stabilised* concepts, procedures and resources for *entire* federation *relative* to background

***UoD* for a significant sustainable sub-community *relative* to common core**



Universes of Discourse

***UoD* for a significant sustainable sub-sub-community**

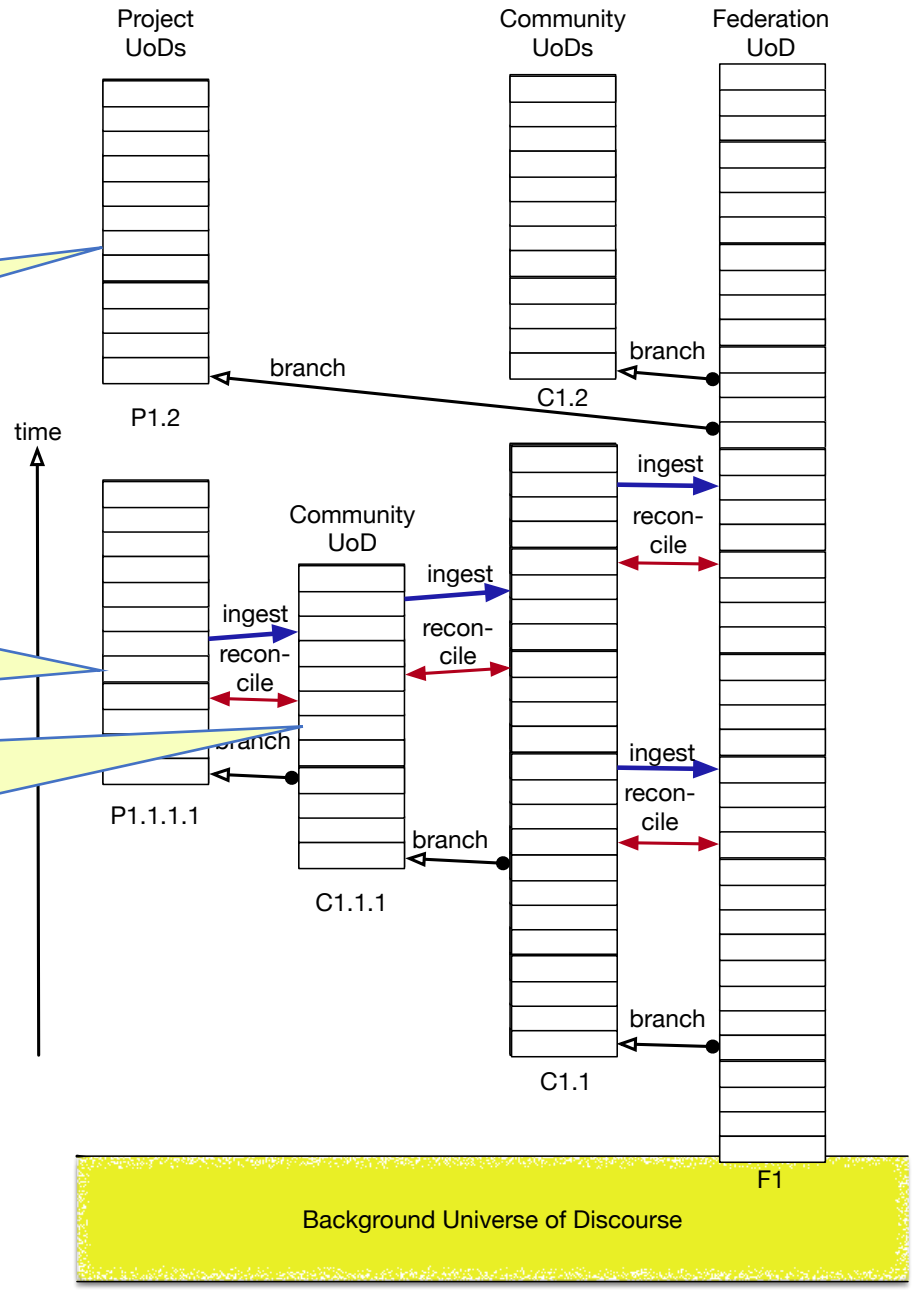


Universes of Discourse

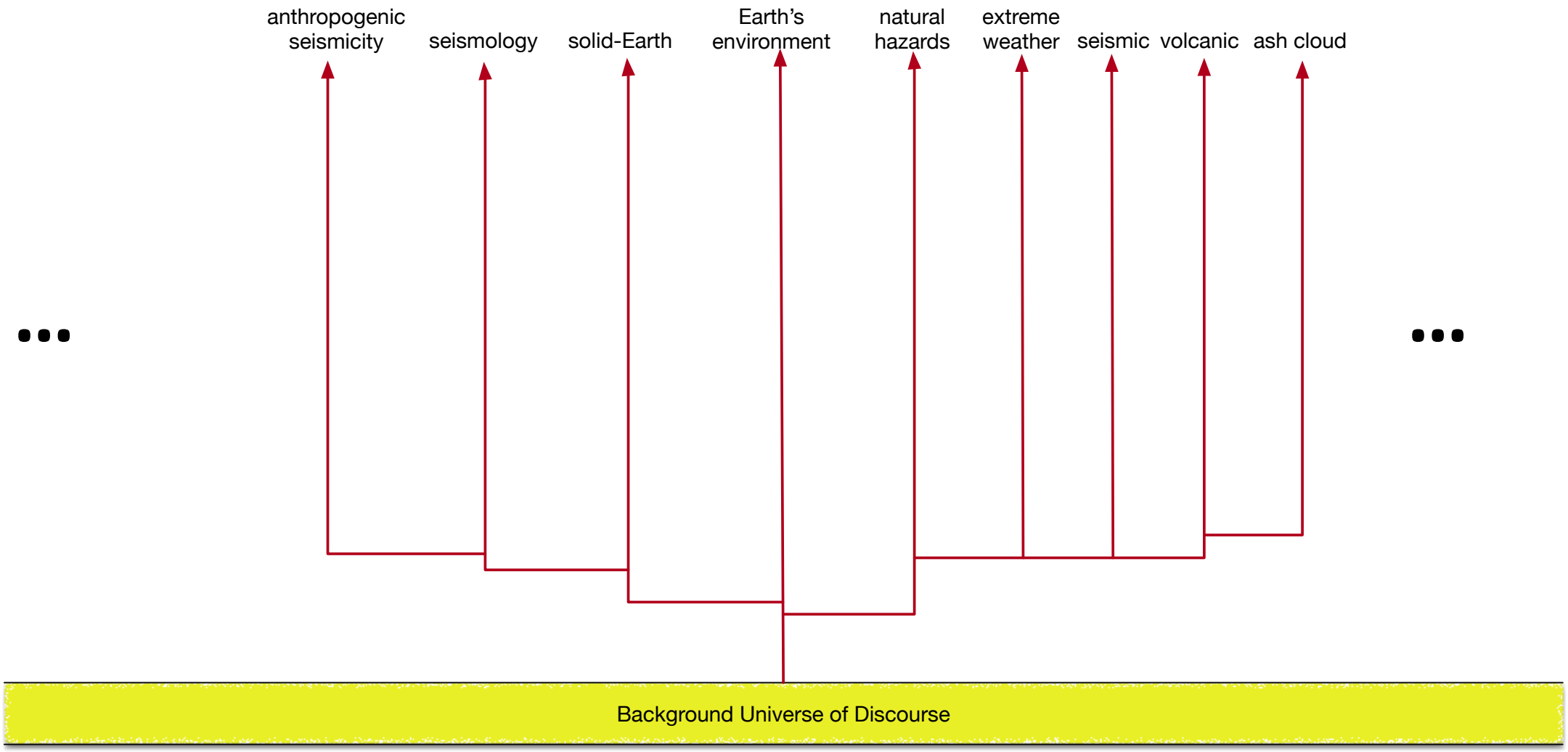
Short **project with unfettered innovation**

Short **project with unfettered innovation still with supporting work environment**

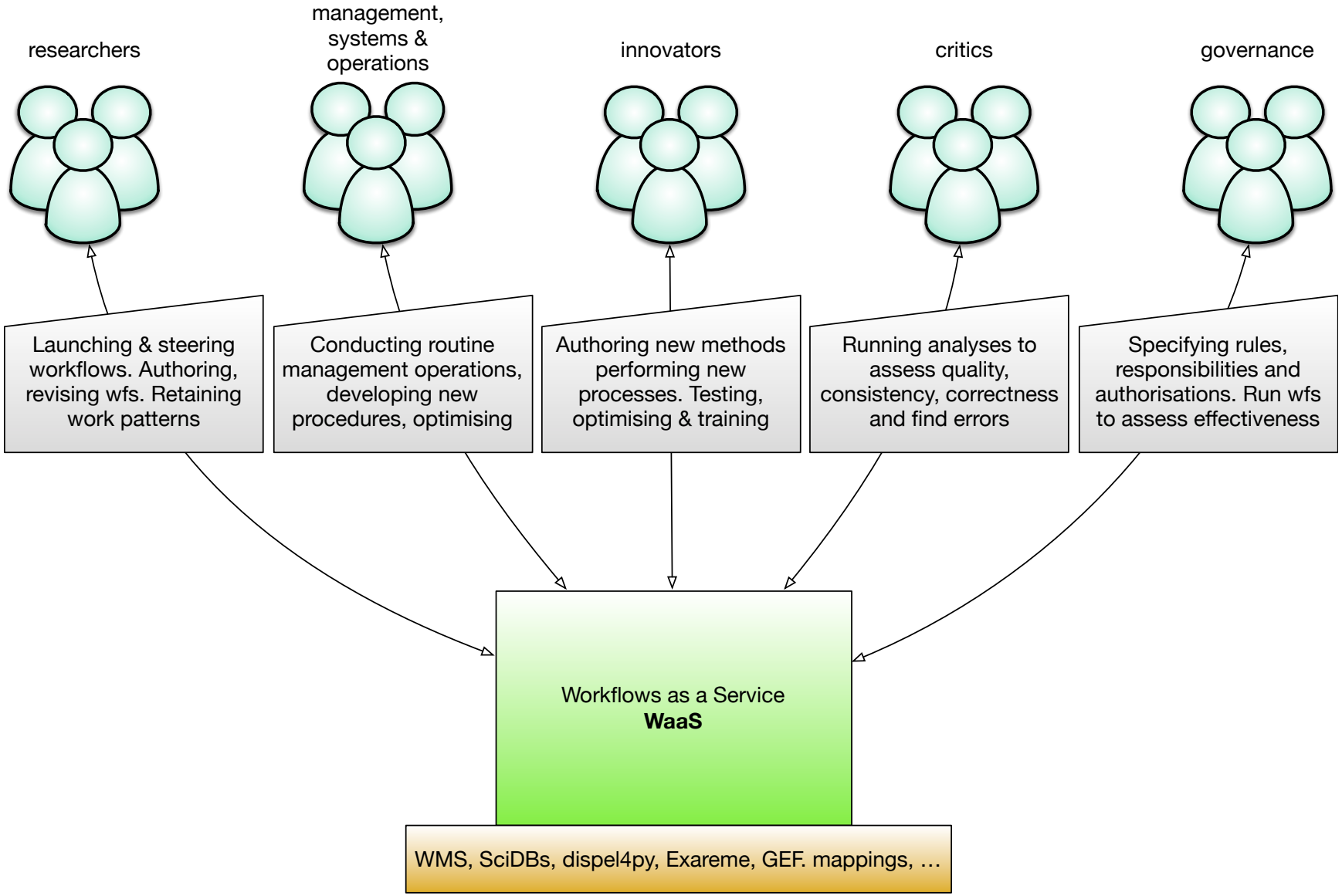
UoD **for a significant sustainable sub-sub-community**



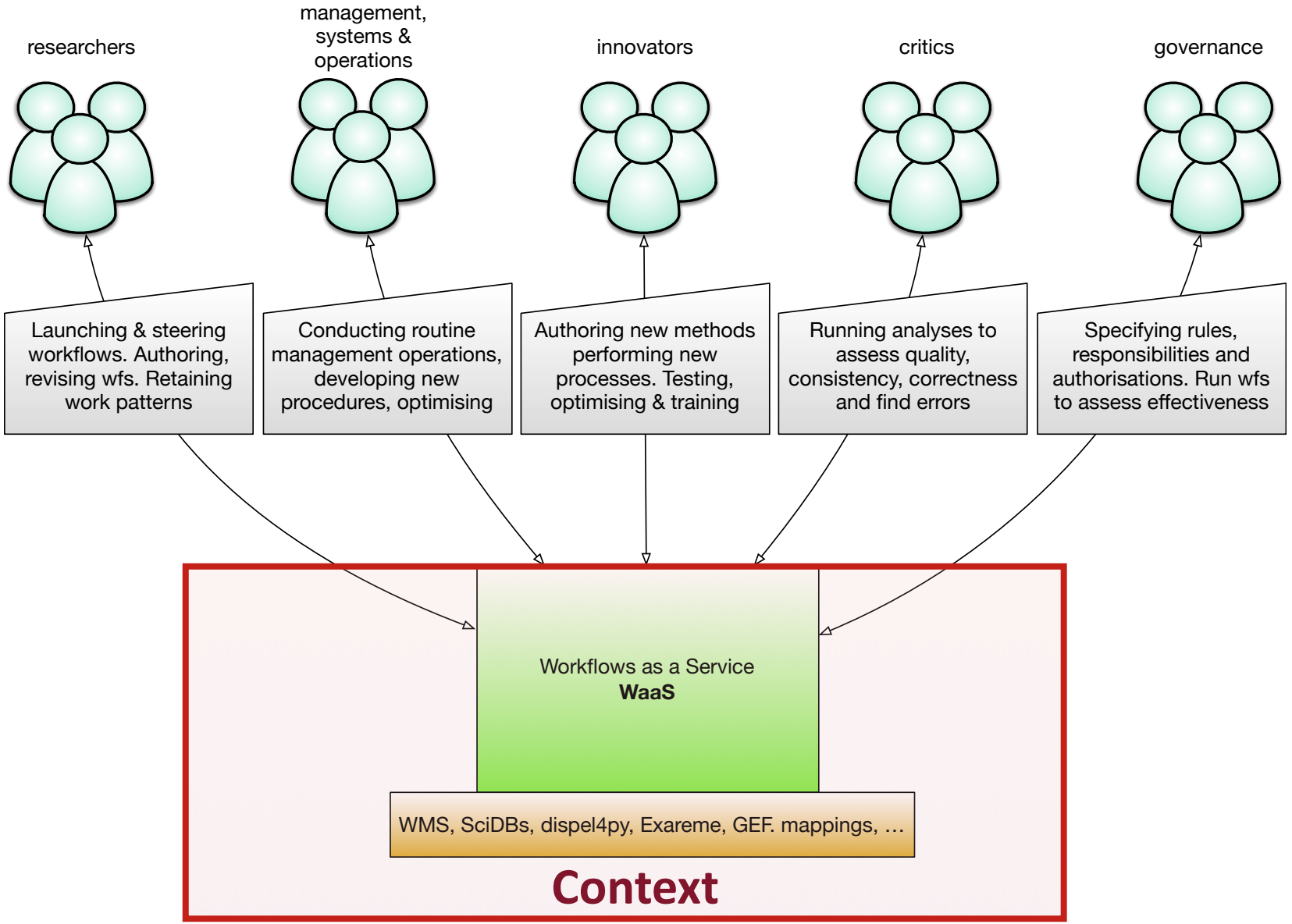
Support social structure



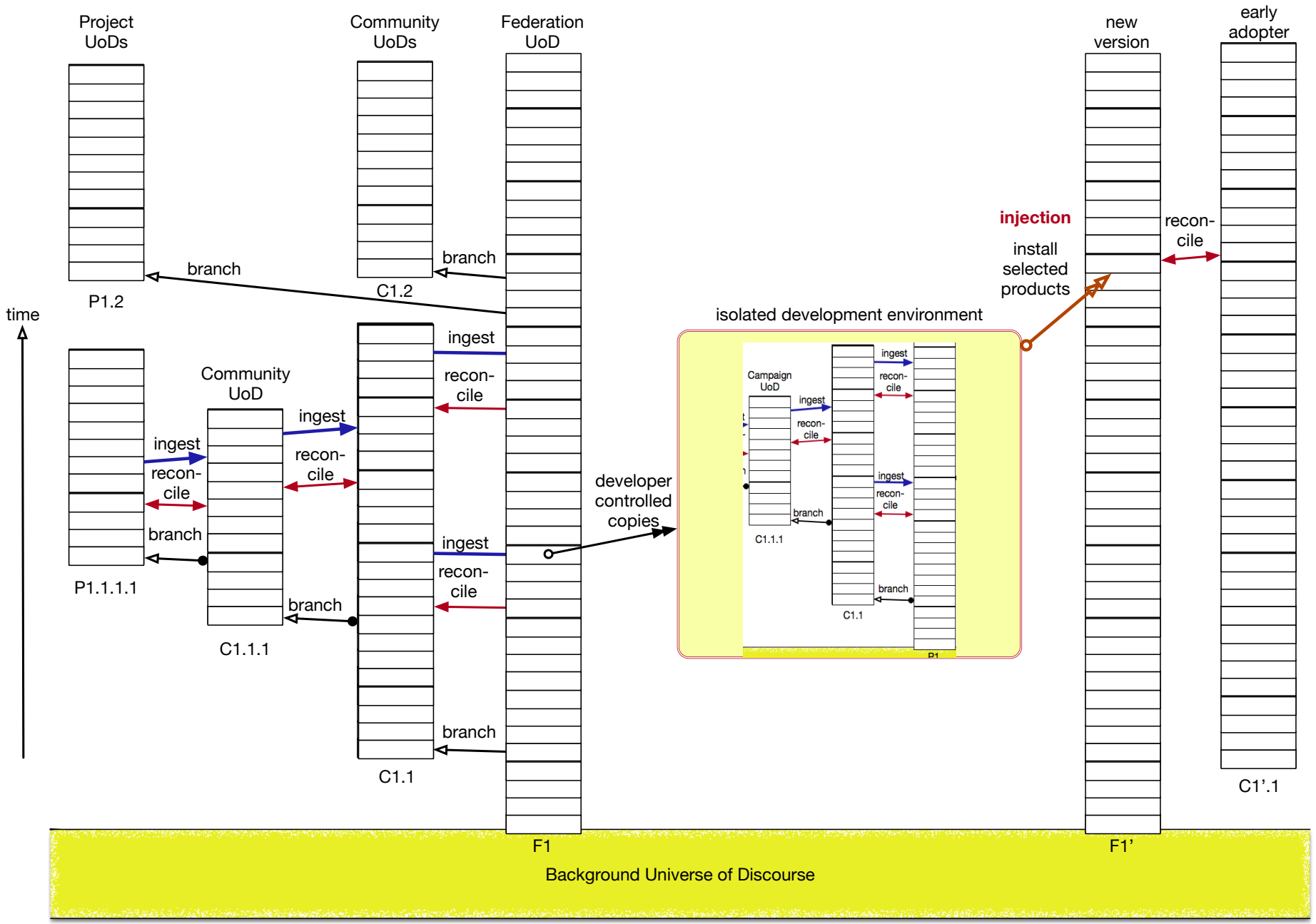
Diversity of working practices



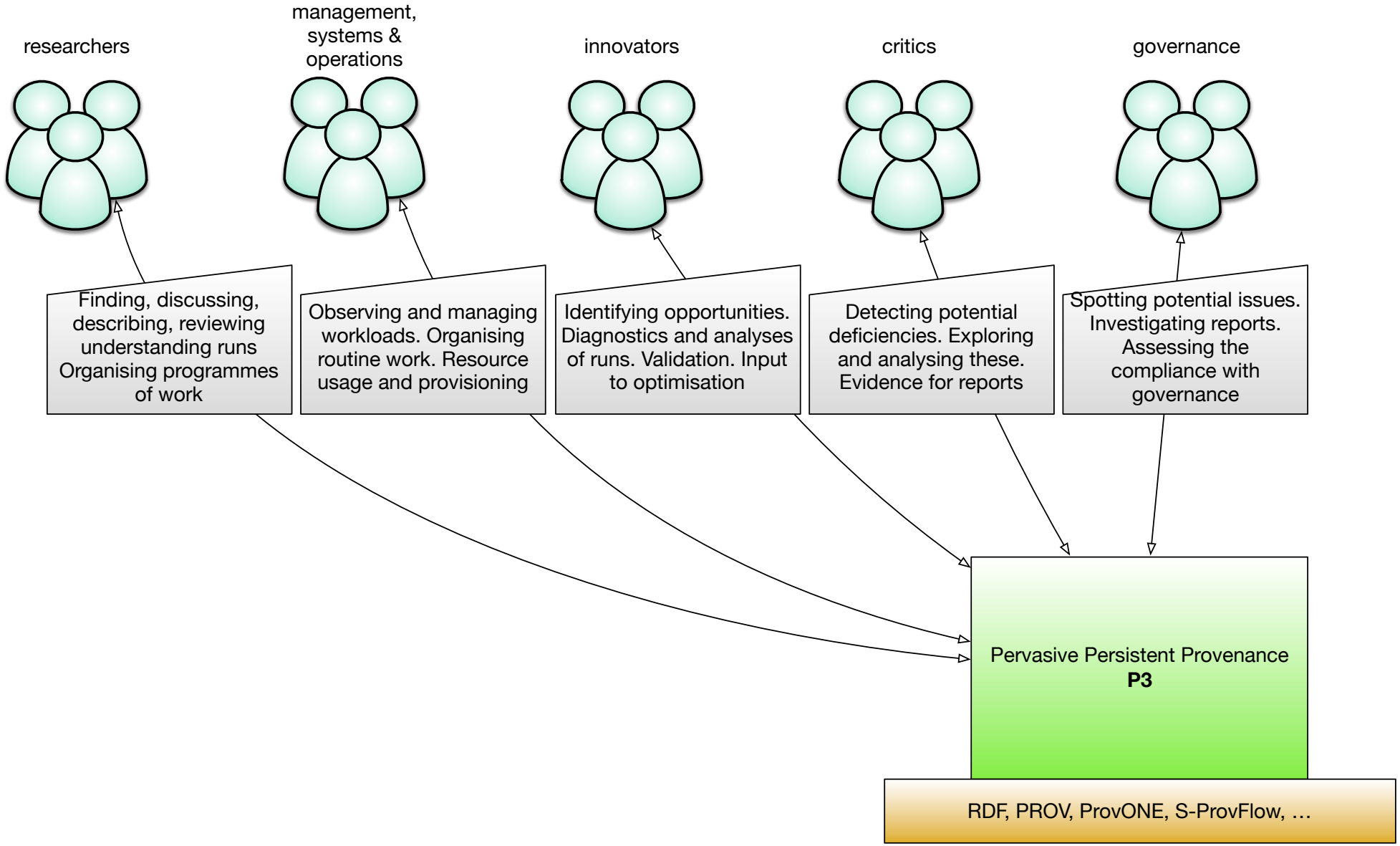
Controls for production + freedom for innovation



Managing change to limit disruption with protection and incremental adoption



Variations through time



Take home messages

- Empower human responsibility and control
- Value diversity and build for it
- Mirror human structures
- Accommodate organisational autonomy
- Nurture innovation while retaining established practices
- Invest in sustaining collaborations and methods
- You can’t dodge complexity our world is complex
- But sustain manageable niches and eliminate technology intrusion

Take home messages

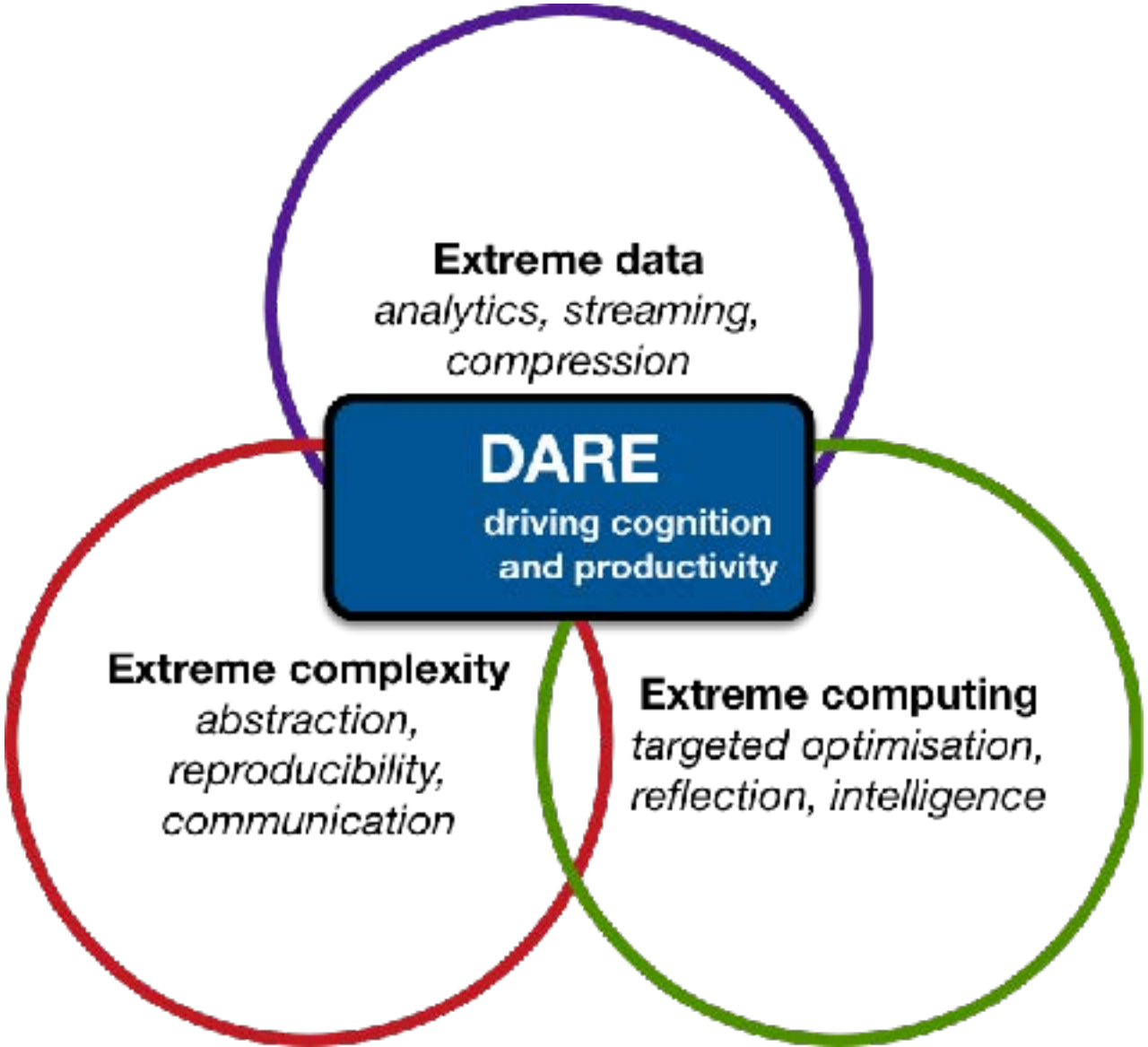
- Empower human responsibility and control
- Value diversity and build for it
- Mirror human structures
- Accommodate organisational autonomy
- Nurture innovation while retaining established practices
- Invest in sustaining collaborations and methods
- You can't dodge complexity our world is complex
- But sustain manageable niches and eliminate technology intrusion

Encourage **innovation and adoption of advances**

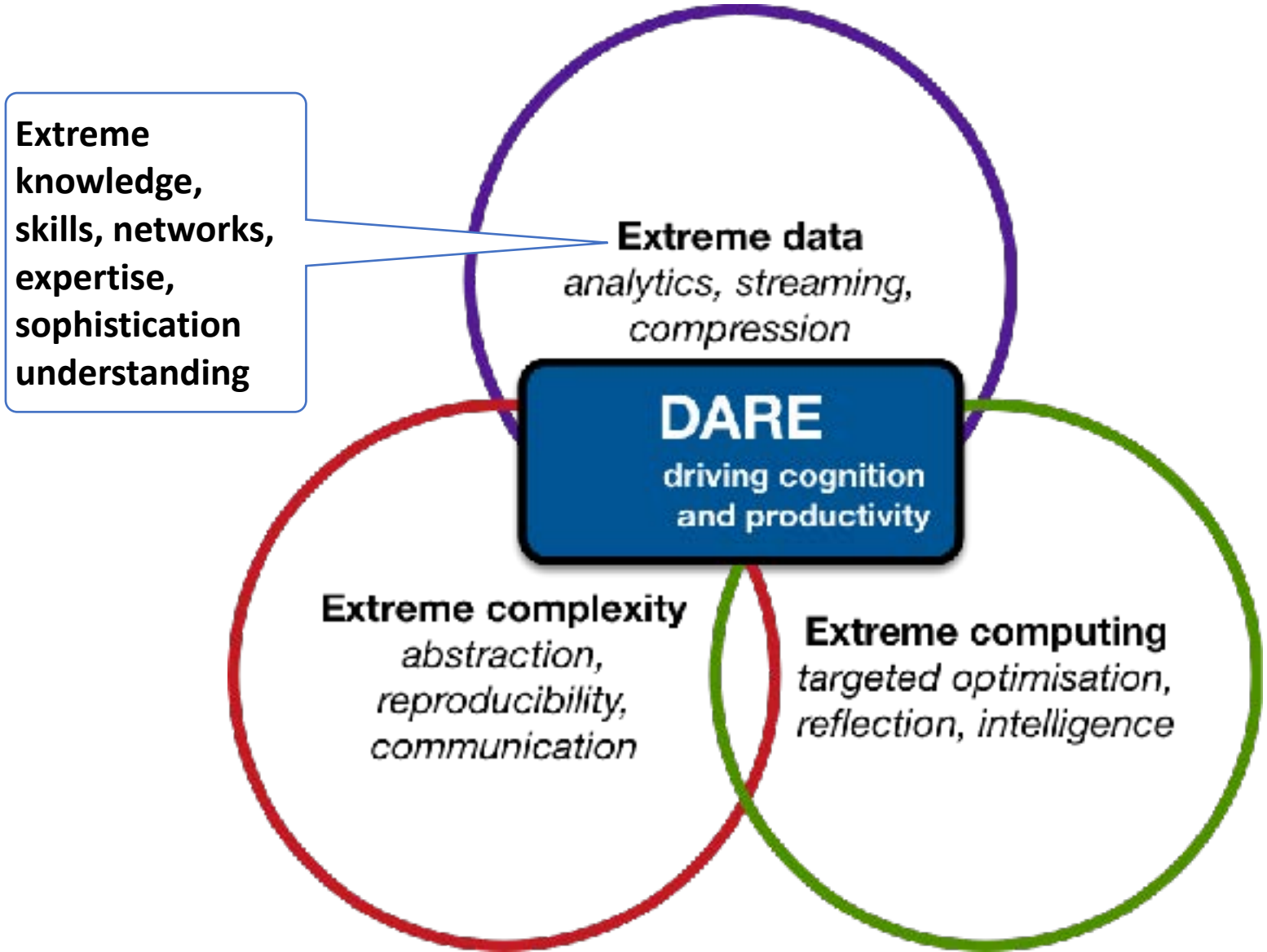
do not

impose **change or uniformity**

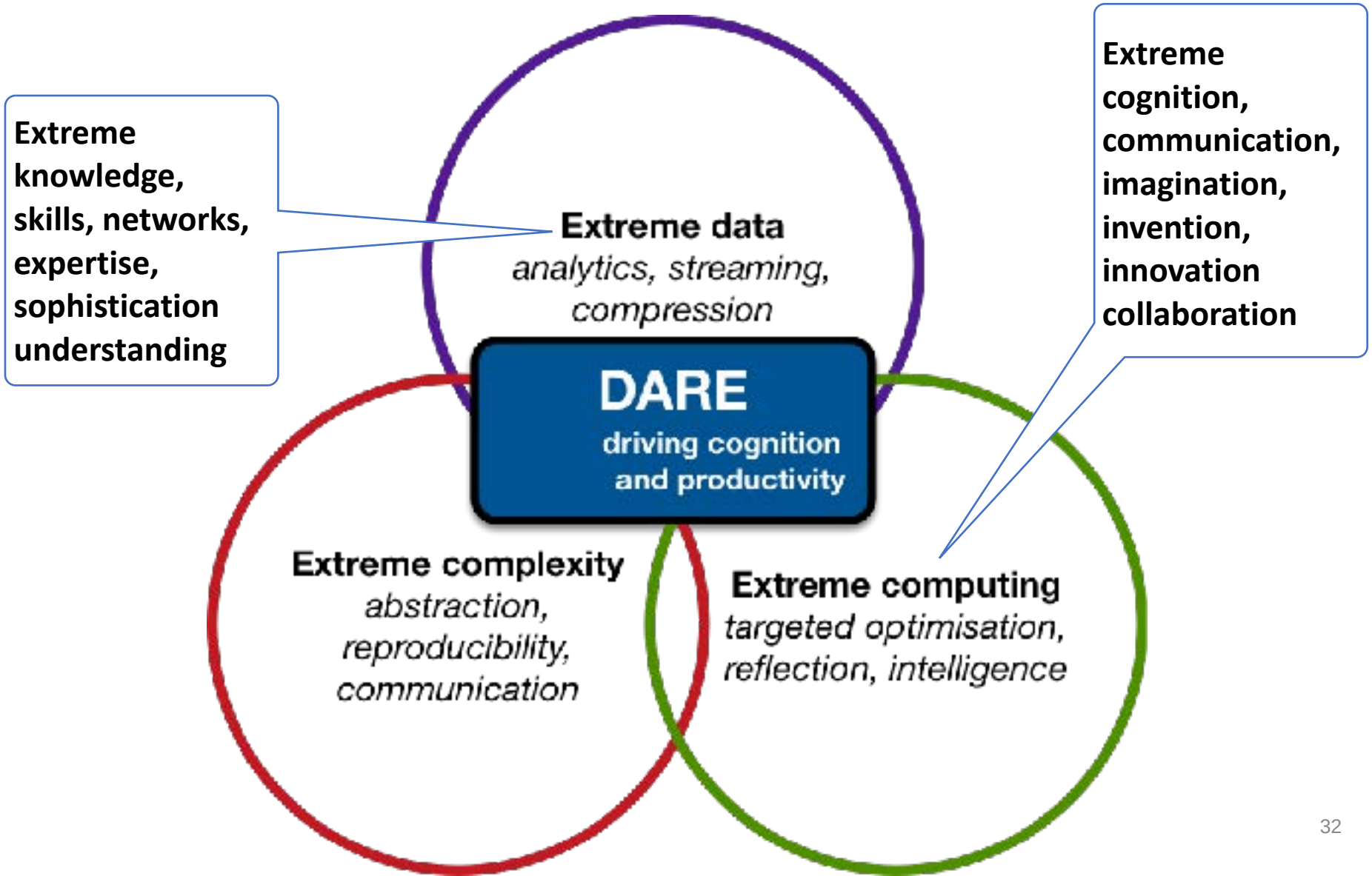
The Human Face of DARE



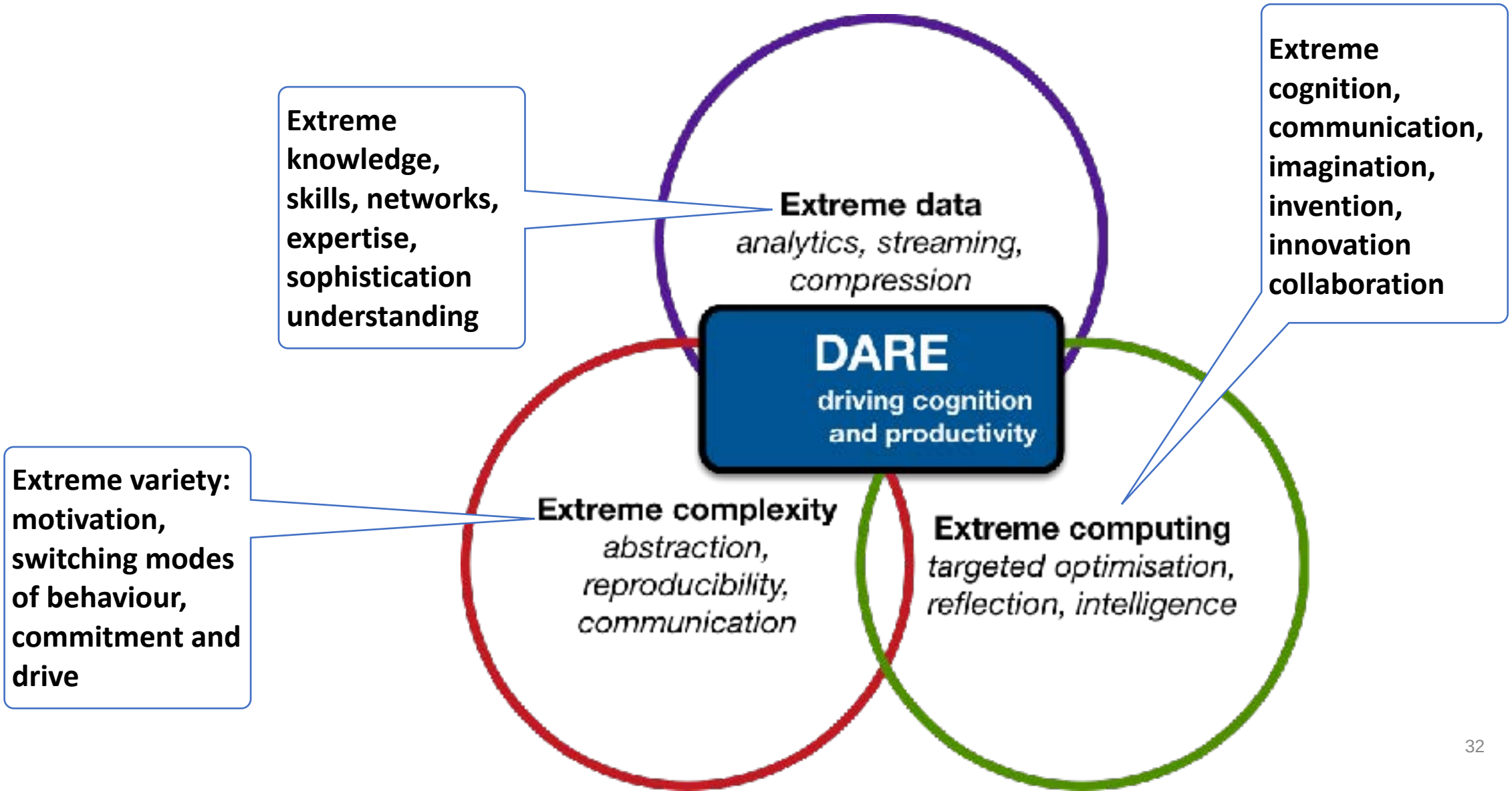
The Human Face of DARE



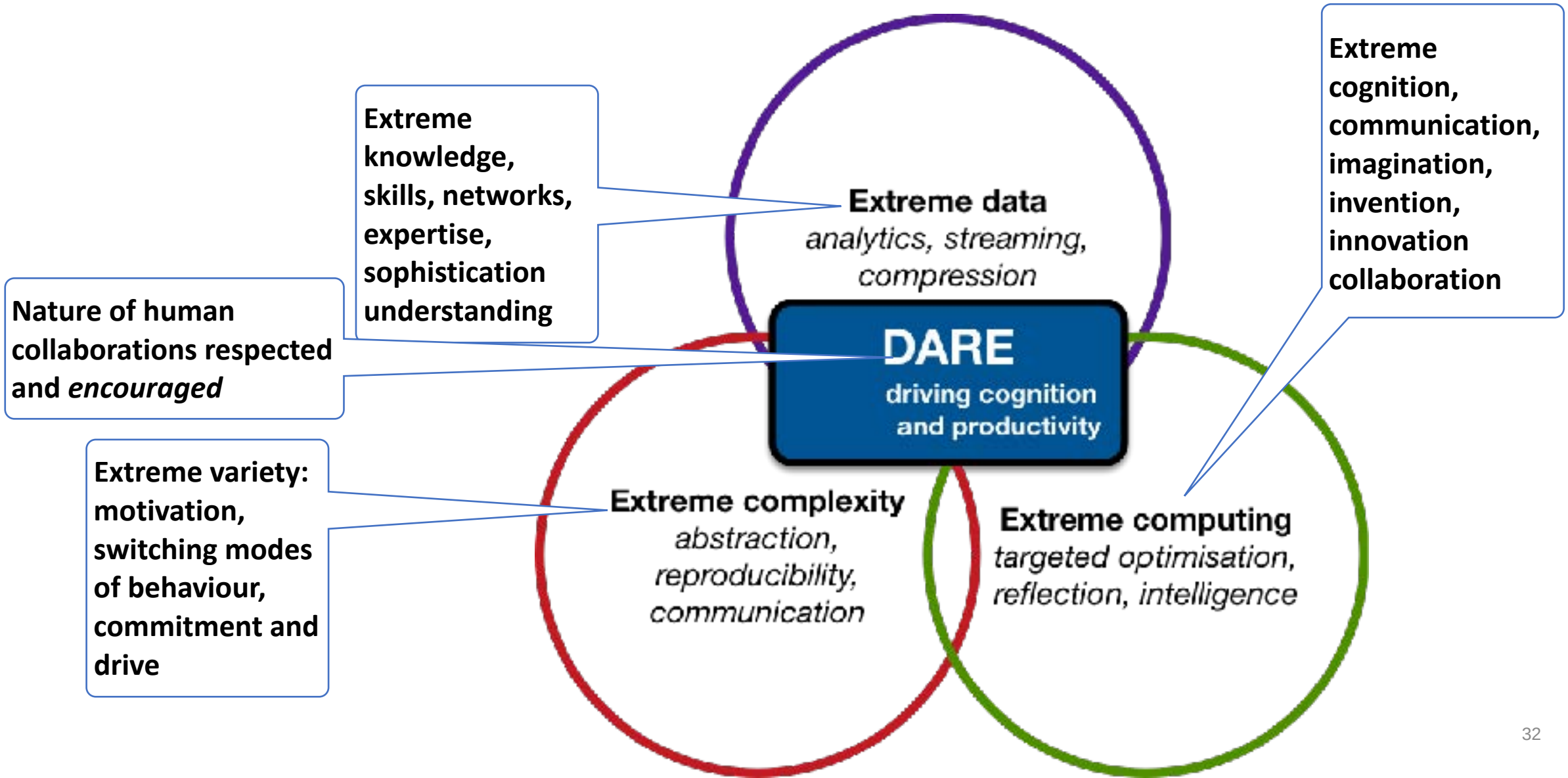
The Human Face of DARE



The Human Face of DARE



The Human Face of DARE



Thank you!

Your questions please

Visit Edinburgh

IWSG 2018

13-15 June 2018

Acknowledgements

Colleagues

Rosa Filgueira, Iraklis Klampanos, Christian Pagé, Alessandro Spinuso, Luca Trani

Projects

UK e-Science, ADMIRE, VERCE, ENVRI, ENVRIplus, SKA-link, DARE

Bibliography

50 years observing Pulsars Jocelyn Bell-Burnell https://www.rse.org.uk/event/fifty-years-of-pulsars-pulasting-radio-stars/

50 years avoiding unnecessary complexity Edsger Dijkstra "*Go To Statement Considered Harmful*", CACM, **11** (3), 147-148, 1968. doi:10.1145/362929.362947

Pegasus' role in gravitational wave detection https://pegasus.isi.edu/tag/ligo/

Accommodating commercial realities http://www.delaat.net/dl4ld/index.html

M.P. Atkinson, S. Gesing, J. Montagnat and I. Taylor, *Scientific Workflows: Past, Present and Future*, FGCS, 75, 216–227, 2017.

R. Filgueira, A. Krause, M.P. Atkinson, I. Klampanos and A. Moreno, *dispel4py: A Python Framework for Data-Intensive Scientific Computing*. In Int. J. of HPC Apps, vol. 31, no. 4, pp. 316-334, 2016

C.S. Liew, M. P. Atkinson, M. Galea, T. F. Ang, P. Martin, J. I. van Hemert, *Scientific workflows: Moving across paradigms*, ACM Comput. Surv. 49 (4) (2017) 66:1–66:39. doi:10.1145/3012429.

A. Spinuso. *Active Provenance for Data-Intensive Research*, PhD thesis, School of Informatics, University of Edinburgh, to be submitted, 2018.

L. Trani, M. Koymans, M.P. Atkinson, R. Sleeman and R. Filgueira. *WFCatalog: A catalogue for seismological waveform data*, Computers Geosciences 106; 101 – 108, 2017.

L. Trani, M.P. Atkinson, D. Bailo, R. Paciello and R. Filgueira. *Establishing Core Concepts for Information-Powered Collaborations – pioneered by solid-Earth sciences*, submitted 2017.