Introducing the Community Capability Model Project

Dr Liz Lyon, Director, UKOLN, University of Bath, UK
Associate Director, UK Digital Curation Centre

IDCC11: Workshop 3: Community Capability Model Workshop,
Bristol, 5 December 2011

Microsoft Research

UKOLN is supported by: JISC

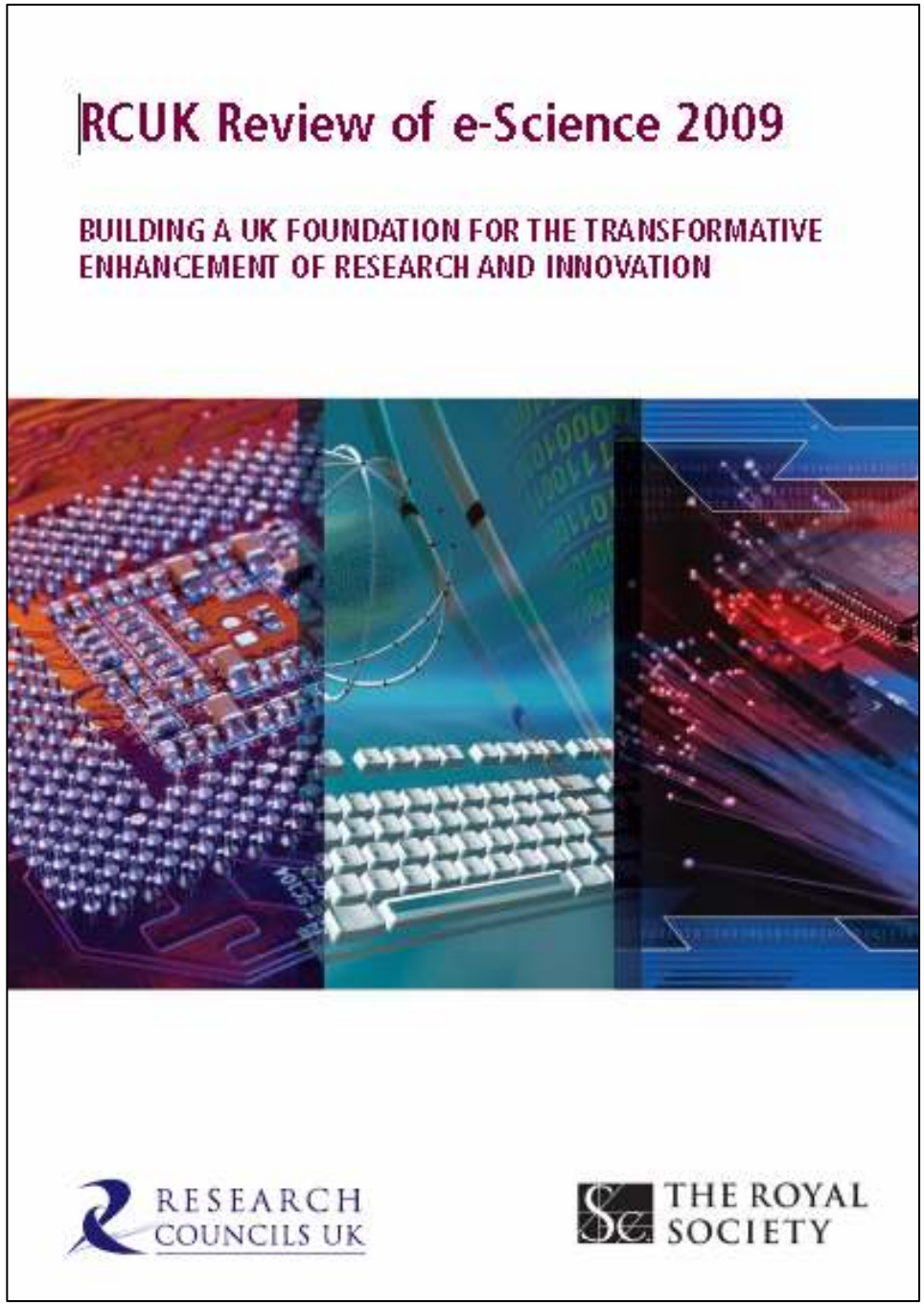
DCC

www.ukoln.ac.uk

A centre of expertise in digital information management

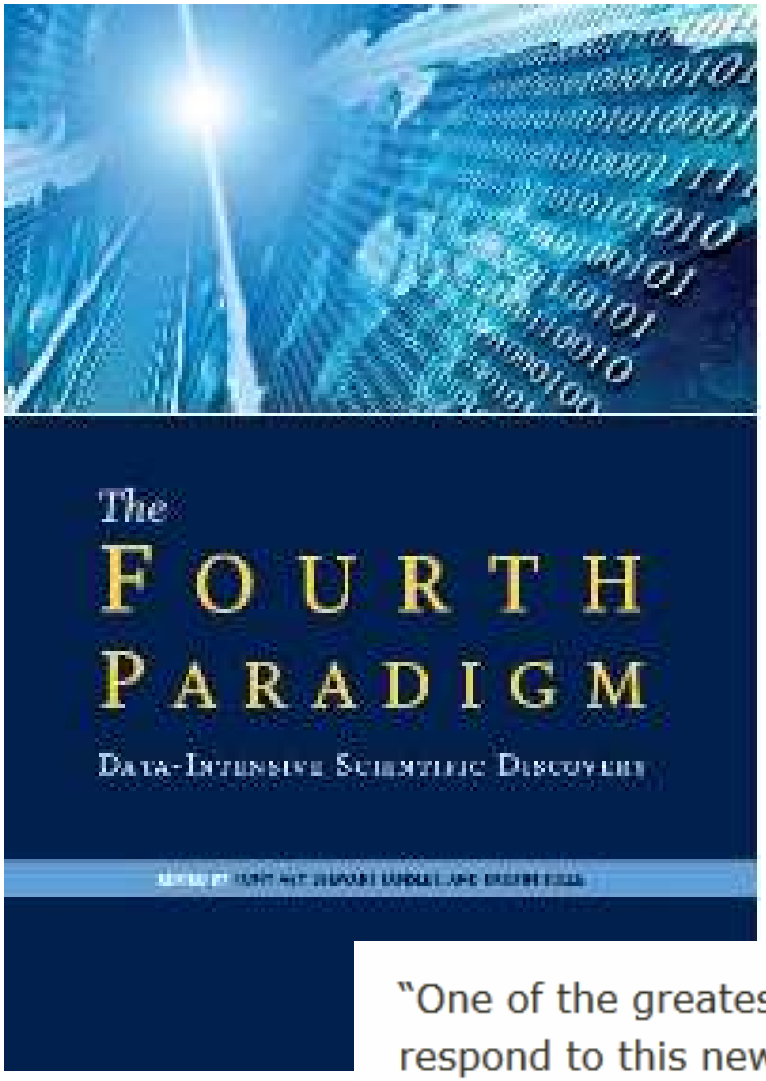
UNIVERSITY OF BATH

THE SUNDAY TIMES UNIVERSITY OF THE YEAR 2011-12



"Data sets are becoming the new instruments of science"

Dan Atkins, Univ Michigan



Microsoft Research

nature.com ▸ nature.com blogs

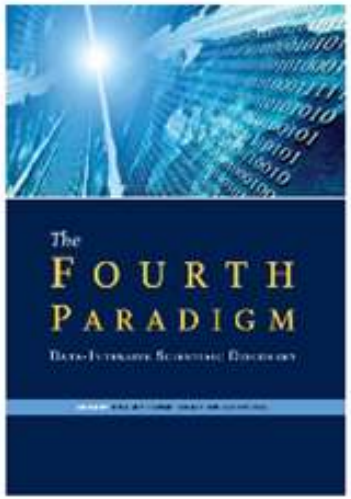
The Fourth Paradigm

a nature network blog by The Fourth Paradi

"One of the greatest challenges for 21st-century science is how we respond to this new era of data-intensive science. This is recognized as a new paradigm beyond experimental and theoretical research and computer simulations of natural phenomena—one that requires new tools, techniques, and ways of working."

— **Douglas Kell**, University of Manchester

Community Capability Model for Data-Intensive Research



Microsoft Research Connections and UKOLN are working in partnership on an exciting new project to develop a Community Capability Model for Data-Intensive Research, building upon the principles described in *The Fourth Paradigm*. This second consultation workshop will focus on discussing and describing scholarly communications to enable data-intensive research, such as collaborative authoring platforms, common data formats and identifiers, data-sharing, data citation and socio-legal issues.

The ultimate aim is to provide a framework that is useful for researchers and funders in modelling a range of disciplinary and community behaviours with respect to the adoption, usage, development and exploitation of cyber-infrastructure for data-intensive research.

UKOLN

UNIVERSITY OF BATH

Microsoft Research Connections

http://communitymodel.sharepoint.com/

What are we trying to achieve?

1. Understand disciplinary and community diversity in data-driven research *(consult)*
2. Unpack the “maturity” concept : identify and deconstruct “capability” factors *(scope)*
3. Explore components and metrics for the capability factors / parameters *(describe)*
4. Develop a Community Capability Model Framework *(model, visualise)*
5. Produce domain mini case studies and business usage cases *(validate)*

Application, value, benefits

- Research Stakeholders
  - PIs, research groups, departments
  - Higher education institutions
  - Research funding agencies
  - Industry, business & innovation partners
- “Getting research done”
- Inform planning and assist decision-making
- Validate funding allocations
- Maximise funder investments
- Accelerate knowledge transfer between domains and across sectors

- York: UK e-Science All Hands Meeting 2011
- Harvard: 2011 eScience Workshop: Transforming Scholarly Communication
- **Bristol: 7th International Digital Curation Conference 5 Dec**
- Stockholm: 7th IEEE eScience Conference 5 Dec
- Australia 2012 10 February, Monash University tbc
- Washington DC 2012 tbc

2011 Workshop programme (consult)

http://www.flickr.com/photos/hantastico/3330775062/

Some definitions & interpretations

- **Capability**: “power or ability to do something, capacity to be used or developed, a facility”
- **Maturity**: “fully grown, fully-developed”
- **Behaviours**: mass adoption & shared usage, community consensus & trust, advanced development & exploitation, embedded skills
- View as a **Capability Spectrum?**
- Norms? Extremes? Trends?
- Components? Taxonomy? Visualisations?
- Indicators, benchmarks, metrics?

Parameters / factors (scope)

Panton Principles
Principles for Open Data in Science

OPEN KNOWLEDGE
OPEN DATA
OPEN CONTENT
OPEN SERVICE

Independent Working – Collaborative Working

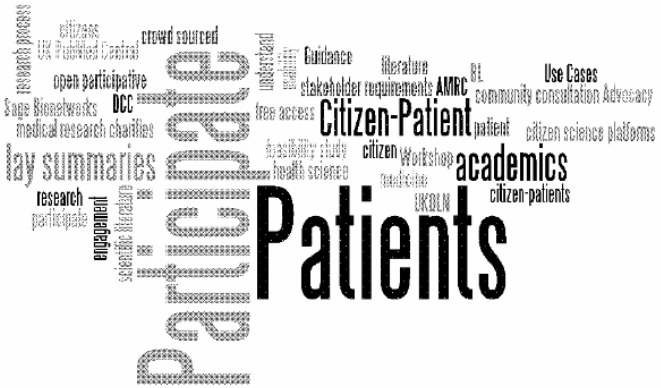
Only individuals and small teams	**Same-discipline, same-sector collaboration and interactions**	*Cross-institutional consortia*
None	**Collaboration and interaction across disciplines**	*Formal collaboration between research groups from different disciplines*
None	**Collaboration and interaction across sectors** *(HE, FE, healthcare, industry…)*	*Joint working on common interests*
None *Basic*	**Public/citizen engagement**	*Dedicated programmes* *Crowd sourcing*

Closed Research – Open Research

No sharing *No details released*	**Openness in the course of research**	*Sharing publicly on the Web* *Full details disclosed*
No sharing *No details released*	**Openness of published literature**	*Open-access publications* *Full details disclosed*
No sharing *No details released*	**Openness of data**	*Sharing publicly on the Web* *Full details disclosed*
No sharing *No details released*	**Openness of methodologies/workflows**	*Sharing publicly on the Web* *Non-standard scripts, tools and software released*
Only own data used	**Re-use of existing data and research**	*Data published in reusable forms* *Multiple existing datasets often combined*

GALAXY ZOO

Patients Participate!



Inclusive or exclusive?

scienceforcitizens.net

1000 Genomes Project Releases Data from Pilot Projects on Path to Providing Database for 2,500 Human Genomes

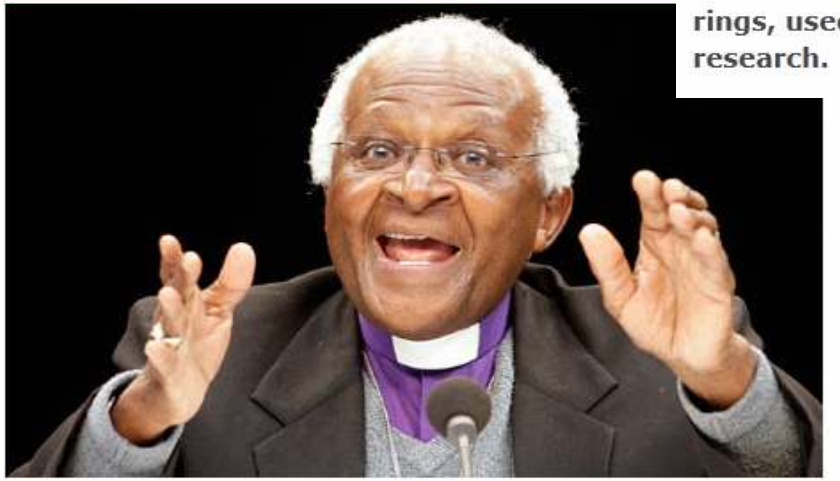
Freely available data supporting next generation of human genetic research

Open or closed?

Desmond Tutu's genome sequenced as part of genetic diversity study

Archbishop Desmond Tutu has had his genome sequer research to reveal the true breadth of human genetic di

Ian Sample, science correspondent
guardian.co.uk, Wednesday 17 February 2010 18.02 GMT
Article history



University told to hand over tree ring data

Queen's University in Belfast has been told by the Information Commissioner to hand over 40 years of research data on tree rings, used for climate research.

Queen's University has been told to

BBC Mobile NEWS

Parameters / factors (scope)

Academic Issues

> 6 years between acceptance and publication	**Lead time from research start to outputs**	*< 18 months between acceptance and publication*
> 2 years between acceptance and publication	**Formal publication lead times**	*< 3 months between acceptance and publication*
Highly risk-averse	**Attitudes towards entrepreneurship, innovation and taking risk**	*Highly innovative and experimental*
Reward models for researchers	*What contributions are recognised and rewarded* *How contributions are recognised and rewarded* *Which measurements are used*	

Information and Communications Technology

ICT support and interoperability in the areas of...

1. **Methods and tools**
   a) **Software**
   b) **Libraries**
   c) **Algorithms**
2. **Data management**
   a) **Capture**
   b) **Processing**
   c) **Storage**
   d) **Curation and preservation**
   e) **Discovery and access**
3. **Communication & collaboration**
   a) **Integration** *(e.g. VREs)*
   b) **Representation** *(e.g. WorldWide Telescope)*
   c) **Citizen science**

Parameters / factors (scope)

Standardization

Availability, quality and use of…

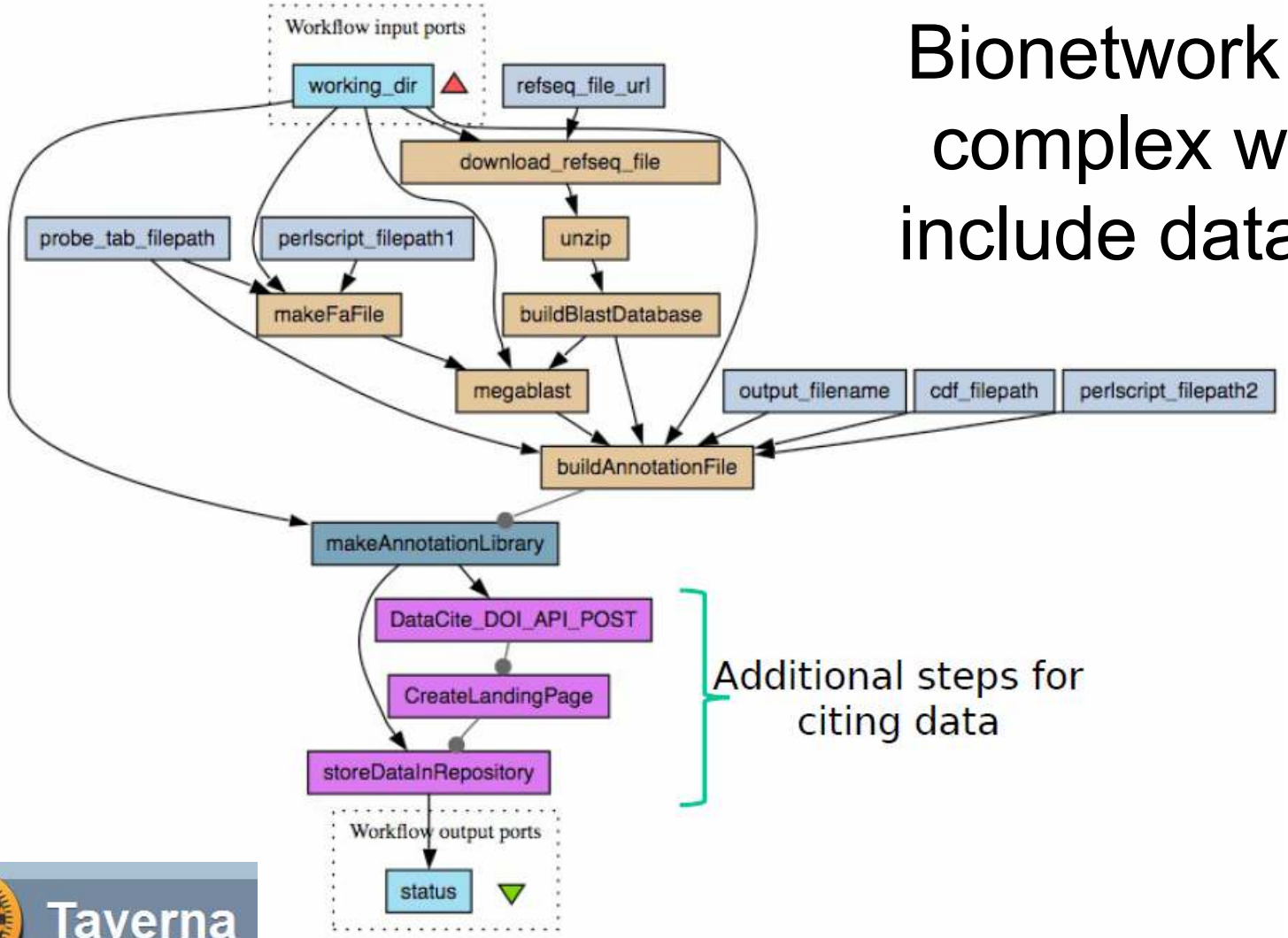
1. **Standard formats**
2. **Standard research methodologies**
3. **Standard processing workflows**
4. **Standard data transfer protocols**
5. **Standard metadata**
6. **Standard vocabularies, semantics & ontologies**
7. **Standard identifiers**
8. **Stable, documented APIs**

Skills and Training

No training opportunities	**Techniques for undertaking data-intensive research**	*Professional qualifications*
No training opportunities	**Technological skills** *(e.g. use of cloud computing)*	*Professional qualifications*
No training opportunities	**Documentation skills** *(e.g. metadata, vocabularies)*	*Professional qualifications*
No training opportunities	**Personal skills** *(e.g. working collaboratively)*	*Professional qualifications*

DCC

Bionetwork models: complex workflows include data citation



Taverna

SageCite

Parameters / factors (scope)

Legal and Ethical Issues

Low awareness of issues	**Approaches to legislative issues** *(e.g. IP, data licensing, rights and patents)*	*De facto, mandated procedures*
Low awareness of issues	**Management of ethical constraints and norms** *(e.g. privacy, confidentiality and consent)*	*De facto, mandated procedures*
None Self-regulation	**Existence of regulatory frameworks**	*Regulation by disciplinary bodies, professional societies*
Lightweight self-review of data	**Approaches to quality control of data & related publications**	*Thorough peer review of data (integrity, appropriateness)*

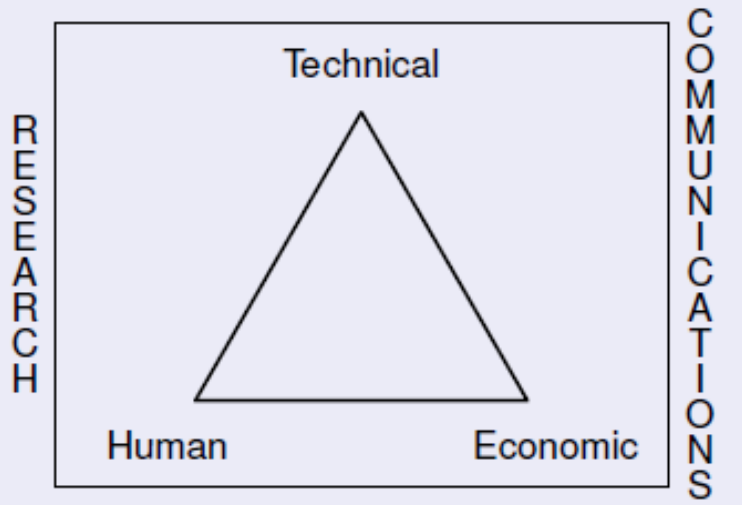
Economic and Business Issues

Economic and business models for funding research		
Economic and business models for sustainability of services *(e.g. storage)*	*Mechanisms for assessing cost/benefits*	
Local funding only	**Globalisation**	*Mainly international funding*
None	**Extent of industrial partnerships**	*Some wholly industrially-funded research*

An Alternative Model?

Alternative Model

Proposed by delegates at the CCMDIR Harvard Workshop



Human

1. Rewards and incentives
2. Legal and ethical issues
3. Independent or collaborative working
4. Closed or open research
   - 4.1 Norms of sharing
   - 4.2 Shared vocabulary
   - 4.3 Shared axioms or speculations
   - 4.4 What communities are formed
5. Skills and training
   - 5.1 Mentoring
   - 5.2 Social value
   - 5.3 . . . in data science
   - 5.4 . . . in vocabularies
   - 5.5 . . . in tools

Economic

1. Operational or start-up funding
2. Long-term sustainability
3. Transactional costs

Technical

1. ICT
2. Platforms
   - 2.1 Database versioning
   - 2.2 Data triage
   - 2.3 Provenance
   - 2.4 Public DMPs
   - 2.5 Public methodologies
3. Access and exposure
   - 3.1 Ease of finding data
   - 3.2 Search/filter facilities
   - 3.3 Publishing open data
   - 3.4 Description of data
   - 3.5 Time dependencies
   - 3.6 Scholarly record of data
   - 3.7 Intra-, inter- and trans-discipline dynamics
4. Standards
   - 4.1 Alternatives to QA metrics tradition
   - 4.2 Shared vocabularies, ontologies
   - 4.3 ‘Minimum information’ standards

Synergy with other approaches

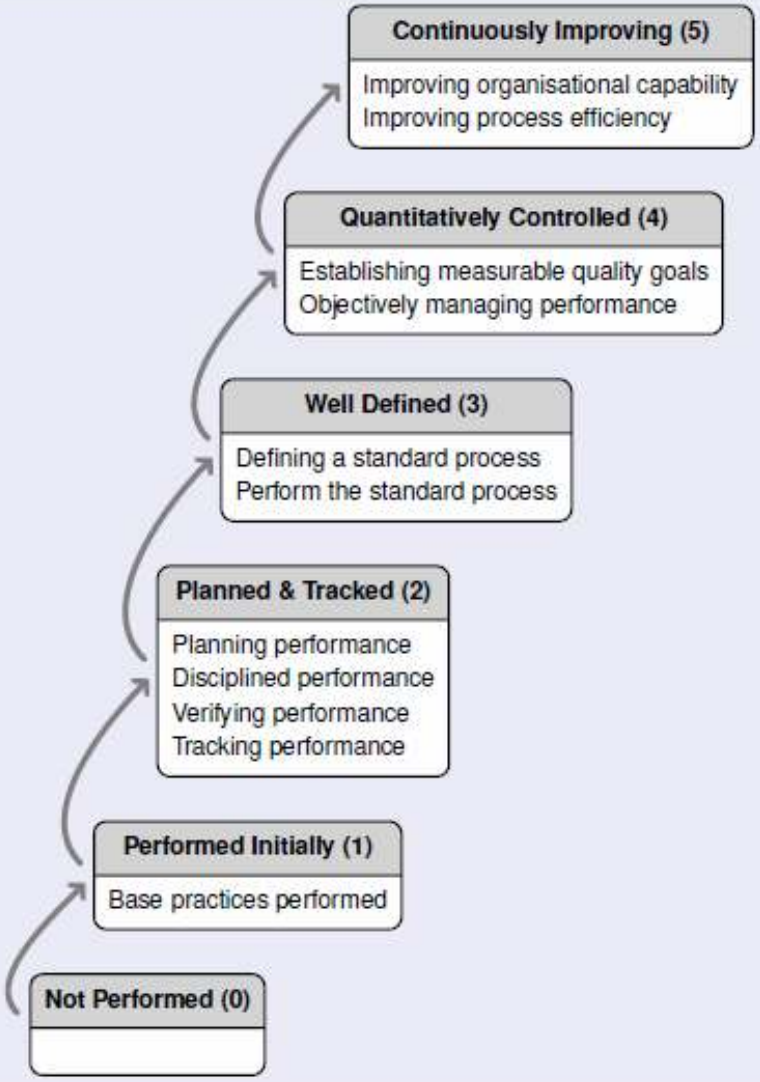
Metrics and measures (describe)

Community Maturity Model

	Strong Hierarchy	Emergent Community	Community	Networked
Strategy	Familiarize & Listen	Participate	Build	Networked
Leadership	Command & Control	Consensus	Collaborative	Distributed
Culture	Reactive	Contributive	Emergent	Activist
Community Management	None	Informal	Explicit Roles & Processes	Integrated Roles & Processes
Content & Programming	Formal & Structured	Some user generated	Community created content & events	Integration of formal & UGC Content
Policies & Governance	No Guidelines for UGC	Restrictive guidelines for UGC	Discrete guidelines for UGC	Comprehensive guidelines for UGC
Tools	Consumer tools used by individuals	Mostly consumer & self-service tools	Mix of consumer & enterprise tools	'Social' functionality is integrated
Metrics & Measurement	Anecdotal	Basic Activities	Activities & Content	Integrated with core business metrics

- The rows represent the eight competencies necessary for successful community management.
- The columns represent the four stages of community maturity, from highly hierarchical organizations to organizations with a networked business ecosystem approach.

Systems Engineering Capability Maturity Model



- The enterprise is divided into *process areas* (e.g. Ensure Quality, Manage Risk).
- Achieving a *capability level* within a process area means implementing a certain set of practices.
- These practices are grouped into *common features* (see figure).
- At Level 1, each process area has its own set of *base practices*.
- At Levels 2–5, all process areas share sets of *generic practices*.

Bate, R., Kuhn, D., Wells, C., Armitage, J., Clark, G., Cusick, K., . . . Reichner, A. (1995). *A systems engineering capability maturity model, version 1.1* (CMU/SEI-95-MM-003). Pittsburgh, PA: Carnegie Mellon University. http://handle.dtic.mil/100.2/ADA303318

Metrics and measures (describe)

Cornell Maturity Model – AIDA – CARDIO

Three-Legged Stool: Metrics of Maturity

Organisation	Technology	Resources
1. Ownership and management	1. Technological infrastructure	1. Financial sustainability plan
2. Policies and procedures	2. Appropriate technologies	2. Review of business plan
3. Policy review	3. Ensuring availability and integrity	3. Technological resources allocation
4. Sharing of/access to research data	4. Integrity of information	4. Risk analysis
5. Preservation and continuity of research data	5. Obsolescence	5. Transparency and auditability
6. Internal audit of research activities	6. Changes to critical processes	6. Sustainability of funding for research data
7. Monitoring and feedback of publication	7. Security of environment	7. Staff skills
8. Metadata management	8. Security mechanisms	8. Staff numbers
9. Legal compliance	9. Implementation of disaster recovery plan	9. Staff development
10. IPR and rights management	10. Metadata creation	
11. Disaster planning and continuity of research	11. Institutional repository	

Levels of Maturity

Acknowledge/ No Action → **Act/ Localised** → **Consolidate/ Co-operate** → **Unify Department/ Internal Integration** → **Externalise/ Embed**

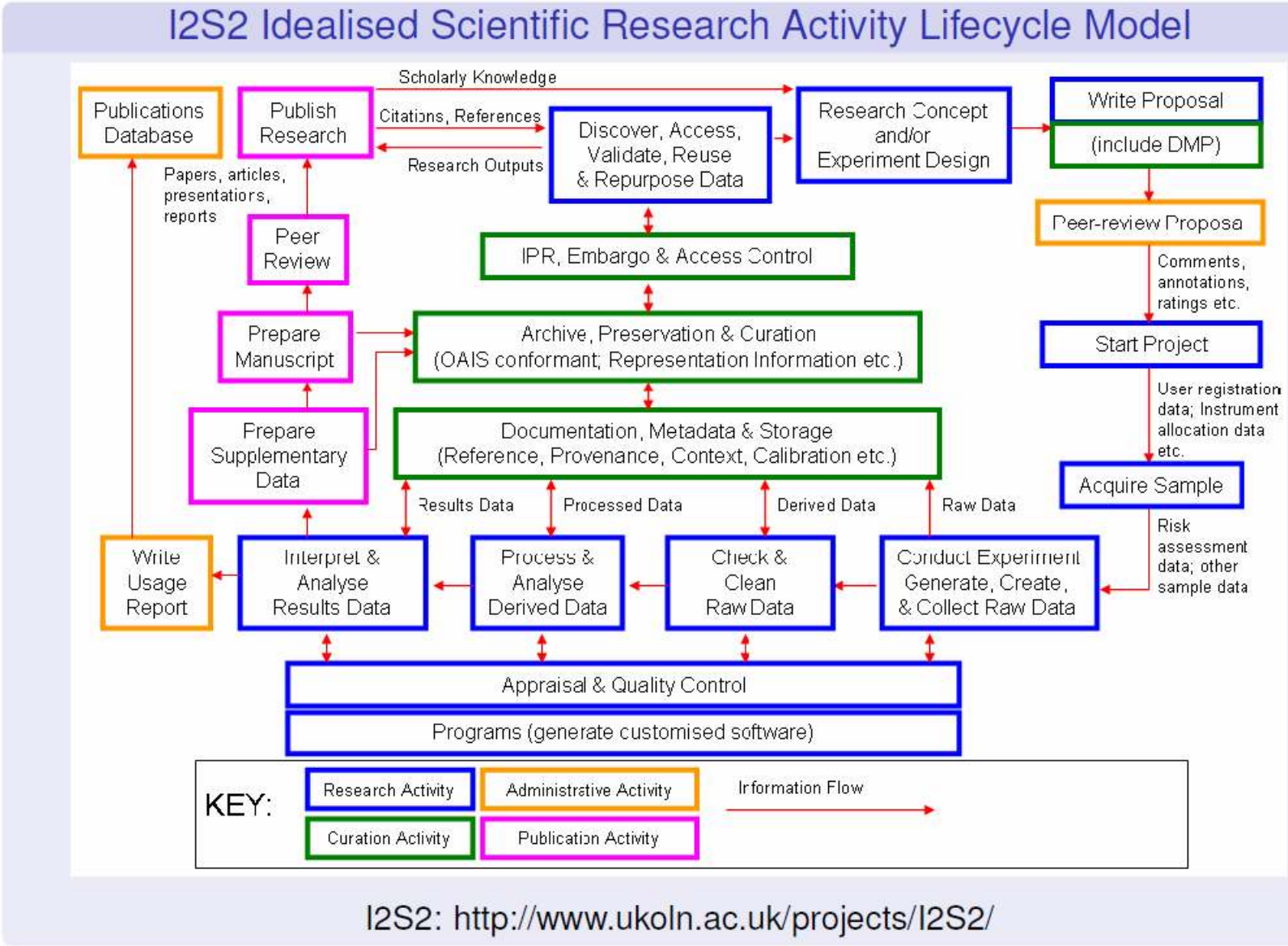
AIDA: http://aida.jiscinvolve.org/wp/

CARDIO: http://cardio.dcc.ac.uk/

CARDIO
collaborative assessment of research data infrastructure and objectives

DCC

Cornell University



Metrics and measures (describe)

A Capability Maturity Model for Scientific Data Management

Kevin Crowston
School of Information Studies
Syracuse University, Syracuse, NY 13244
crowston@syr.edu

Jian Qin
School of Information Studies
Syracuse University, Syracuse, NY 13244
jqin@syr.edu

Proceedings of the American Society for Information Science and Technology
Volume 47, Issue 1, pages 1–2, November/December 2010

Exemplar

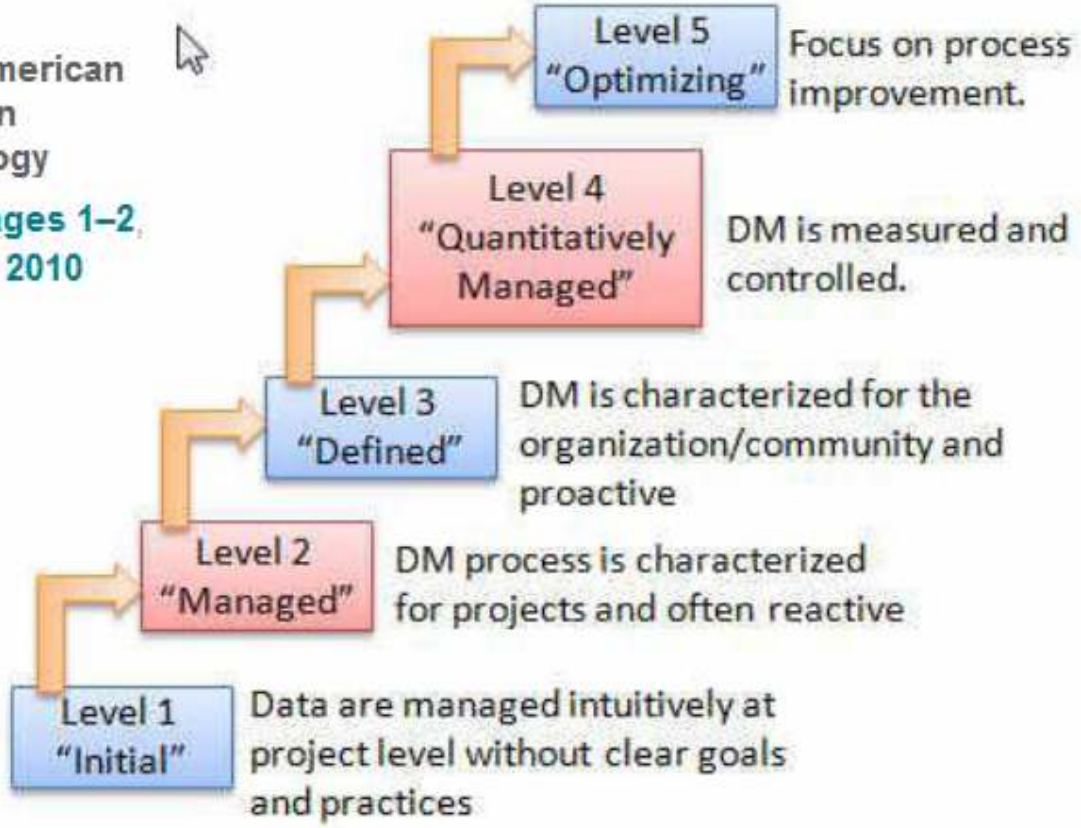
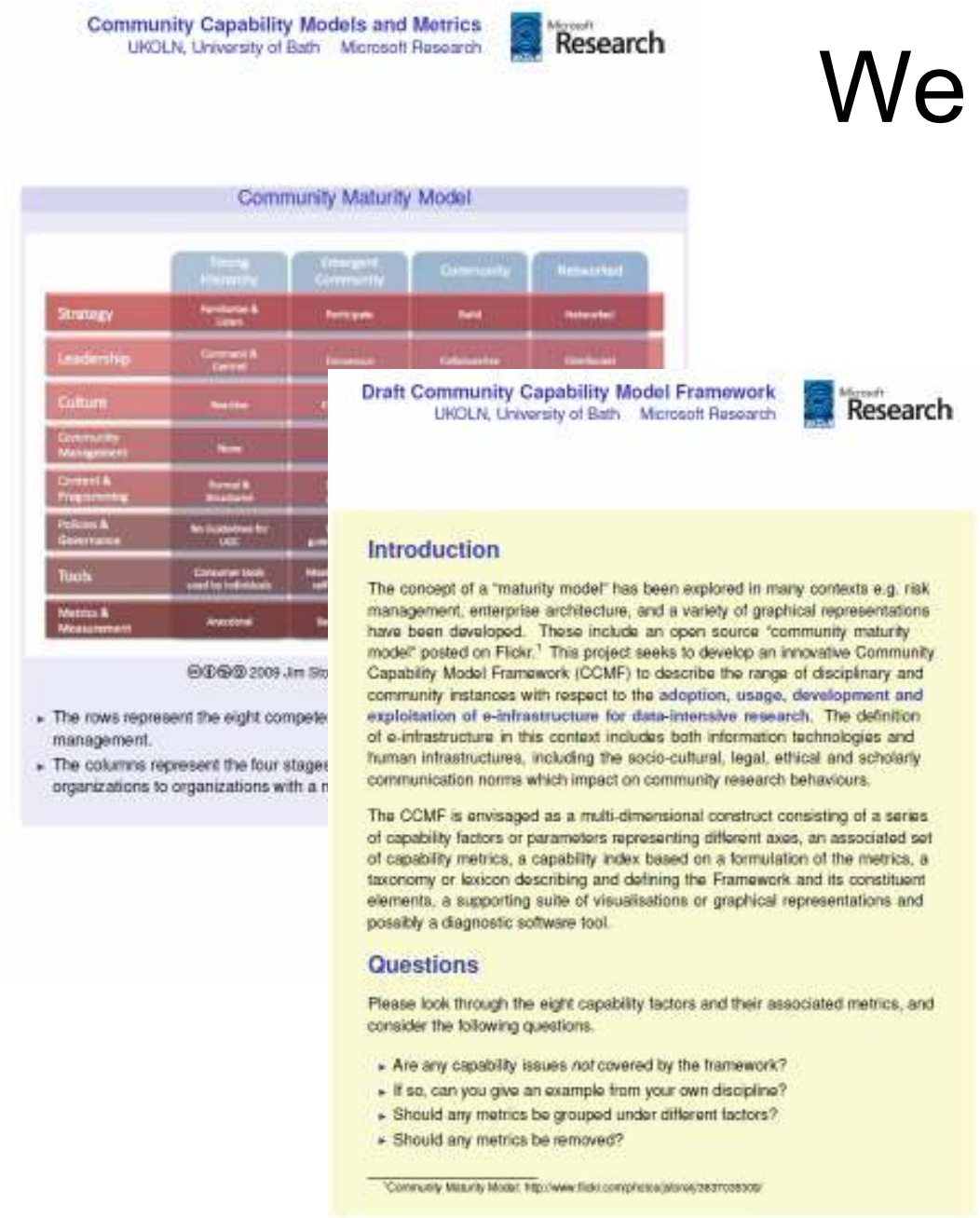


Figure 1. Capability maturity levels for SDM.

We need your help!



Community Capability Models and Metrics
UKOLN, University of Bath Microsoft Research

Community Maturity Model

Strategy				
Leadership				
Culture				
Community Management				
Content & Programming				
Policies & Governance				
Tools				
Metrics & Measurement				

©©©© 2009 Jim Sto

- The rows represent the eight compete management.
- The columns represent the four stage organizations to organizations with a

Draft Community Capability Model Framework
UKOLN, University of Bath Microsoft Research

Introduction

The concept of a "maturity model" has been explored in many contexts e.g. risk management, enterprise architecture, and a variety of graphical representations have been developed. These include an open source "community maturity model" posted on Flickr.[1] This project seeks to develop an innovative Community Capability Model Framework (CCMF) to describe the range of disciplinary and community instances with respect to the adoption, usage, development and exploitation of e-infrastructure for data-intensive research. The definition of e-infrastructure in this context includes both information technologies and human infrastructures, including the socio-cultural, legal, ethical and scholarly communication norms which impact on community research behaviours.

The CCMF is envisaged as a multi-dimensional construct consisting of a series of capability factors or parameters representing different axes, an associated set of capability metrics, a capability index based on a formulation of the metrics, a taxonomy or lexicon describing and defining the Framework and its constituent elements, a supporting suite of visualisations or graphical representations and possibly a diagnostic software tool.

Questions

Please look through the eight capability factors and their associated metrics, and consider the following questions.

- Are any capability issues *not* covered by the framework?
- If so, can you give an example from your own discipline?
- Should any metrics be grouped under different factors?
- Should any metrics be removed?

[1] "Community Maturity Model": http://www.flickr.com/photos/jbum/2837008309/

Groupings and Gaps?
What can the CCM provide to institutions?

Next steps

- 2012 Prepare White Paper describing the CCM Framework for consultation
- Develop case studies (PIs, institutions, funding agencies) and business case
- Australia and Washington DC workshops: (validate) - test the strawman Framework

http://communitymodel.sharepoint.com/