

‘So who’s that new kid on the block?’

Peter Burnhill

Director, Phase One

(on loan from EDINA national data centre)

UK Digital Curation Centre: Timeline

- identified in Report commissioned by JISC Cttee for Support of Research (Lord & Macdonald, May 2003)
 - Twin drivers: Digital Preservation & Continuing Access (e-Science)
- call to set up DCC in JISC Circular 6/03, June 2003
 - Joint funding: JISC (£750K pa for outreach, services & development) and e-Science Core Programme (£250K pa for research)
- 3 year awards made, with March & September 2004 starts
 - Consortium: Universities of Edinburgh (lead), Glasgow & Bath with CCLRC (Rutherford and Daresbury Laboratories)
- preparing for launch in October/November 2004
 - at National eScience Centre

2



Digital | Curation | Centre

Aims & Objectives

‘quality improvement in data curation & digital preservation’

- Initial focus: data as evidence for scholarly conclusions
- Wider remit: worlds of scholarly communication & eLearning
- twin aims: excellence in research & excellence in service
- need to bridge across communities:
 - universities & research institutes
 - scientific data tradition & document tradition
 - multi-sectoral, international

digital curation = f(digital preservation, data curation)

- actions needed to maintain and utilise digital data & research results over entire life-cycle, for current and future use.

f(data curation, digital preservation)

digital curation: ... over their life-cycle, for current & future generations of use ...

- data curation [high current/ongoing interest]
 - actions needed to maintain and utilise digital data & research results over entire life-cycle
 - data creation & management; adding value; generating new sources of information & knowledge, for use
- digital preservation [longevity; fall off in interest]
 - long-run technological/legal accessibility & usability
 - storage, maintenance & accessibility of information content in digital material over the long-term, for use
 - OAIS concept of designated community

Data curation in action

- **Astronomy**

-
- Integrating and analysing distributed data (AstroGrid)
 - publishing multi-TB sky surveys (SuperCOSMOS & WFCAM)
 - interoperability standards (IVO Alliance)

- **BioInformatics**

- data publishing: generic tools for XML export (EBI Biomart)
- annotation tools for massive data sets (Pubmed, VOTable)
- archiving tools for dynamic data sets (biological DBs)

- **Environmental sciences**

- spatio-temporal annotation (OS Mastermap/ Mouse Atlas)

- **Document management**

- Tools for capture & normalisation (Xena)
- Repository certification (RLG Task Force)

Digital preservation approaches

- Migration & Refreshment
- Emulation & Encapsulation
- Digital Archaeology & Rescue
- Document Format Specification

Communities of Practice: Social Sciences (IASSIST)

- History of sharing – economical in terms of both data collector and respondent
- Data about humans – problems of confidentiality confronted early on
- Mixed blessing of agreed proprietary formats (OSIRIS, SPSS, etc.) allows migration
- ‘Future-proofing’ - 30 years of data advocacy!
 - Tradition of data archiving & data citation
 - Building new data standards out of common experience
- data archivists, & data librarians: are they now digital curators?
- www.iassistdata.org

Document format specification

- They borrowed from records management tradition - institutions to create documents in standard or open formats, which are easier to preserve.
- Much easier to do in a strict records management environment with a published policy of retention schedules and a clear knowledge of internally produced records.
- Stipulating a specific file format is harder in a research environment where a wide range of digital materials are produced and have to be preserved.
- *The move to DDI DTD in social science data world may be seen as an example of this preservation technique.*

Unifying Themes for a Digital Curation Centre

- ‘data as evidence’
 - for one or more designated communities
- ‘archival responsibility’
 - at one or more institutional levels
 - with institutional policies & individuals’ competence
- engage/discover communities of practice, to invoke/provoke good practices
 - appraisal & retention/disposal
 - logical & physical integrity: authenticity/security
- locate research problems in productive domains
 - eg Informatics, Law School

Design & Planning for the Long

- we have large territory, and large expectation
 - multi-disciplinary, multi data type, multi tradition/profession
 - national and international, but also local and hidden from view
 - a lot is going on
- how to ensure that we do something sensible with the ££'s and the trust we have been given?
- who/what should we plan to affect/effect?
 - policy-makers; 'responsible curators'; (researchers?)
- how do we wish to be judged, and when?
- collaboration & win-win-win scenarios

focii of attention in set-up phase

- **Users: client, peer and policy communities**
 - outreach & community support; service definition/delivery; development co-ordination; research agenda
 - user requirements analysis: Leona Carpenter (Focus Groups)
- **Consortium: 'organisation' from partner participation**
 - roles; commitment; norming/performing; operational communication; consortium agreement (IPR)
- **Employers: institutional settings**
 - re-deployment/appointments; accommodation; commitment/reporting

11

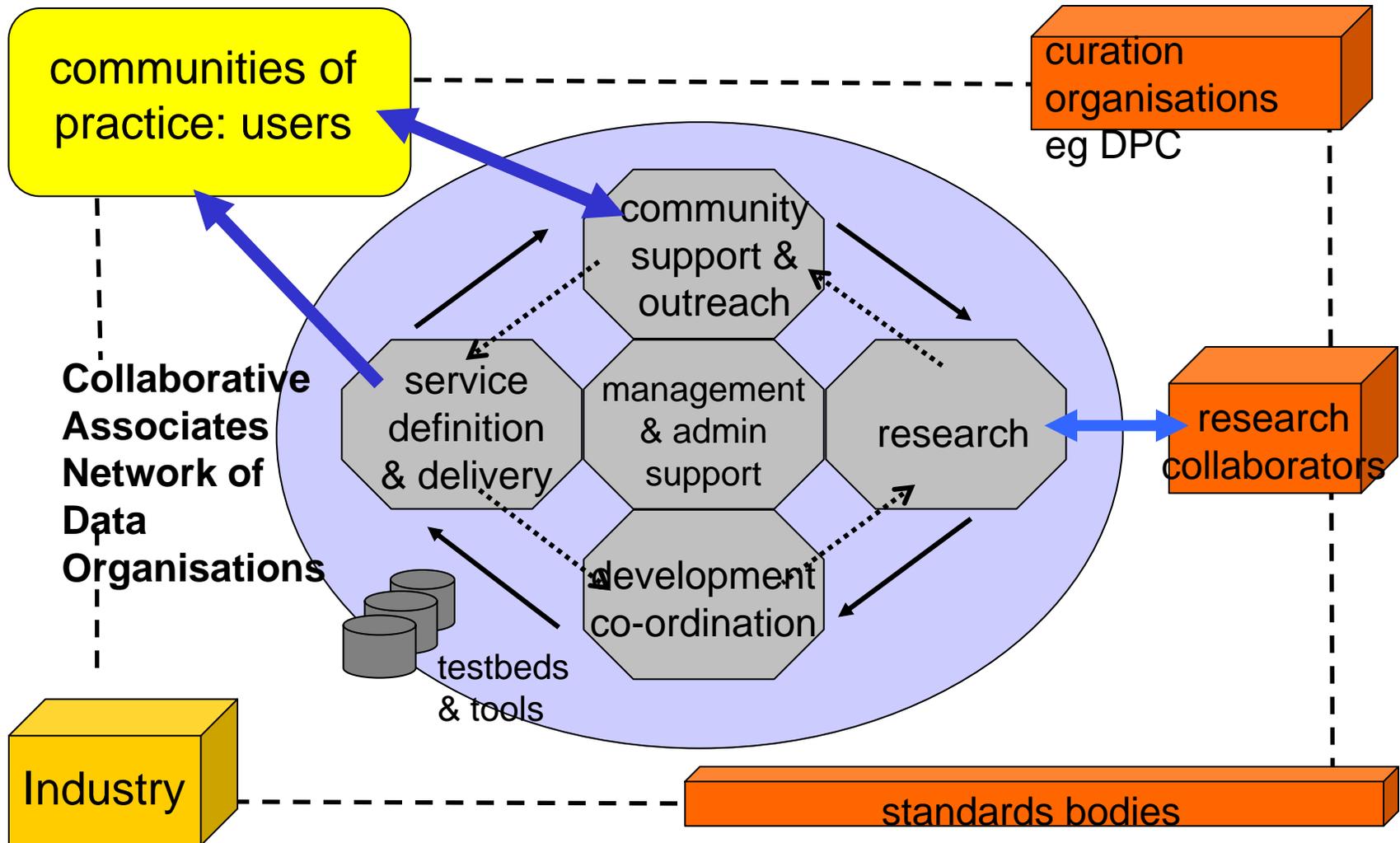
-> Project Plan, as living document

Responsibilities across the DCC

- Them with titles ...
 - Peter Burnhill, Director (Phase One)
with Robin Rice, Phase One Project Co-ordinator
 - EDINA & Data Library, University of Edinburgh
 - Peter Buneman Research Director (& PI on EPSRC grant)
 - Informatics, University of Edinburgh
 - Liz Lyon, Associate Director (Community Support & Outreach)
 - UKOLN, University of Bath
 - Seamus Ross, Associate Director (Service Definition & Delivery)
 - HATII, University of Glasgow
 - David Giaretta, Associate Director (Development)
 - CCLRC
- Two significant & well known 'Ex Portfolio' names
 - Malcolm Atkinson, Director, NeSC
 - Chris Rusbridge, Director, Information Services, UofGlasgow

12

developing the collaborative model



Deliverables

Management & Support	Outreach & Community Support	Service Definition & Delivery	Development Co-ordination	Research Agenda
Project Setup, Project Planning	User Requirements Study	Quality Assurance & Evaluation	Development Monitoring (including new funding opportunities)	Annotation Data integration and publication
Governance structure	Publicity	Advisory Service & Helpdesk		
Collaborative Associates Network of Data Organisations	Web Portal Events	Repository Guidelines	Registry / Repository Development	Appraisal and long-term preservation
Appoint full-time Director	Training & Professional Development Programme	Catalogue of tools Curation manual	Representation Information Development	Socio-economic and legal context
Consortium Agreement	E-Journal	Technical Standards Framework	Testing and Certification Processes	Performance and Optimisation

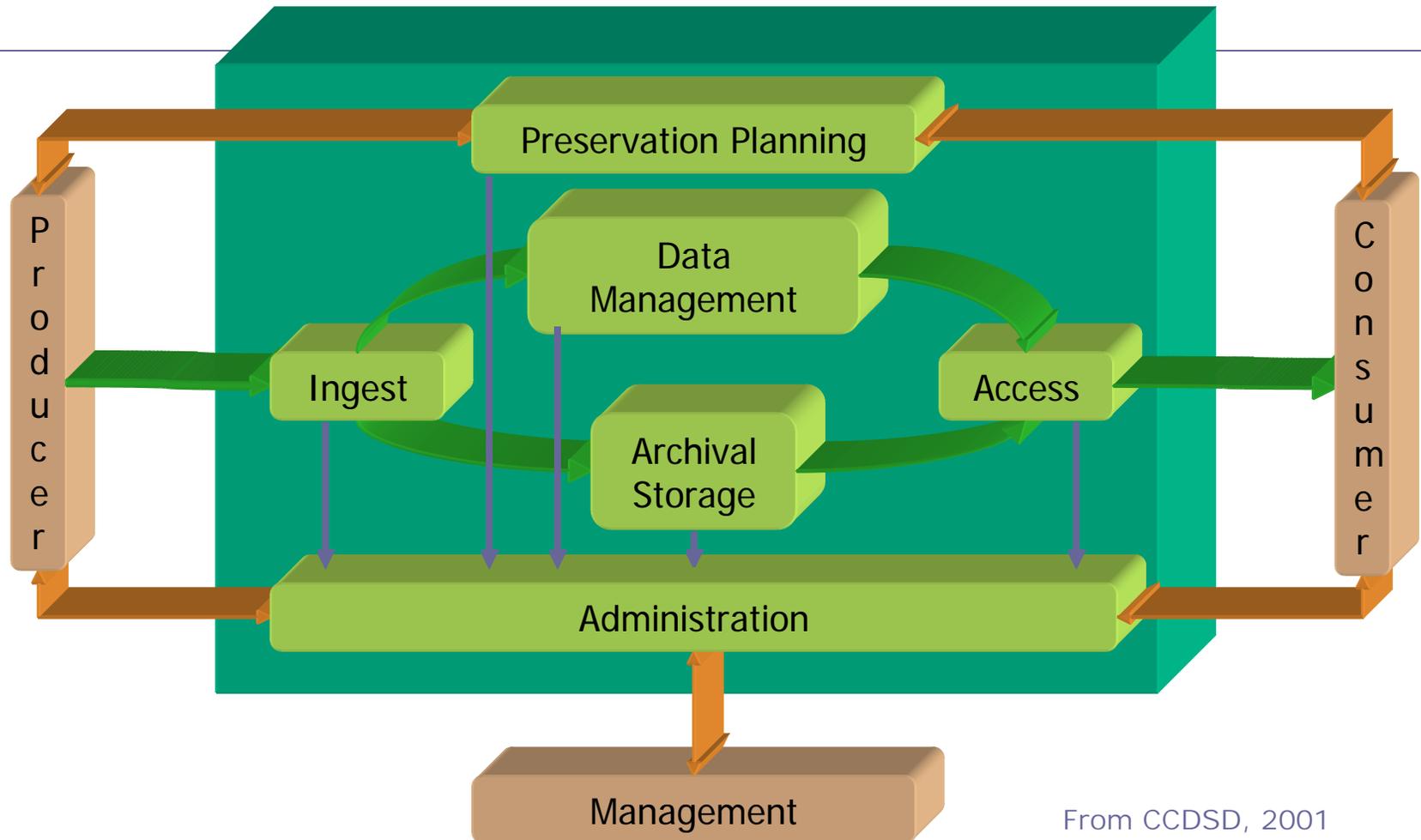
Services & Development

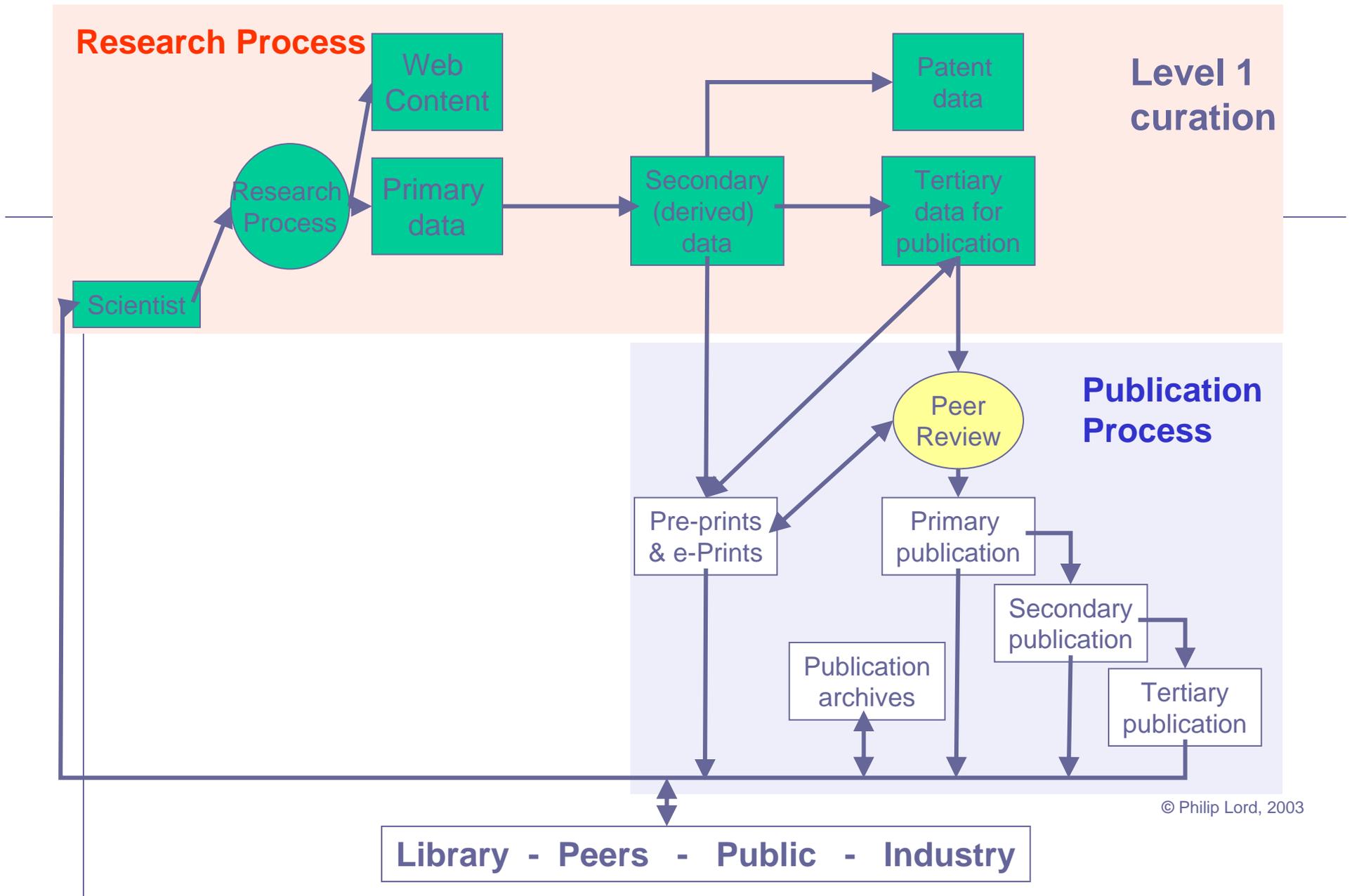
- Turns Research into ‘Products for Research’ that our communities can use with confidence
 - tracking and testing tools and standards
 - that are correct, usable, reliable, well documented
 - e.g. for ingest, repository management, data exchange, ontologies
 - working with tool developers wherever possible
 - developing testbeds & interworking with other testbeds
 - aim to gain leverage formats
 - working with other projects worldwide
 - using generic tools and techniques
 - to develop strategies for emerging digital formats
 - Metadata standards
 - long-term viability of metadata
- Registries underpin this work to provide basis of Advisory Service

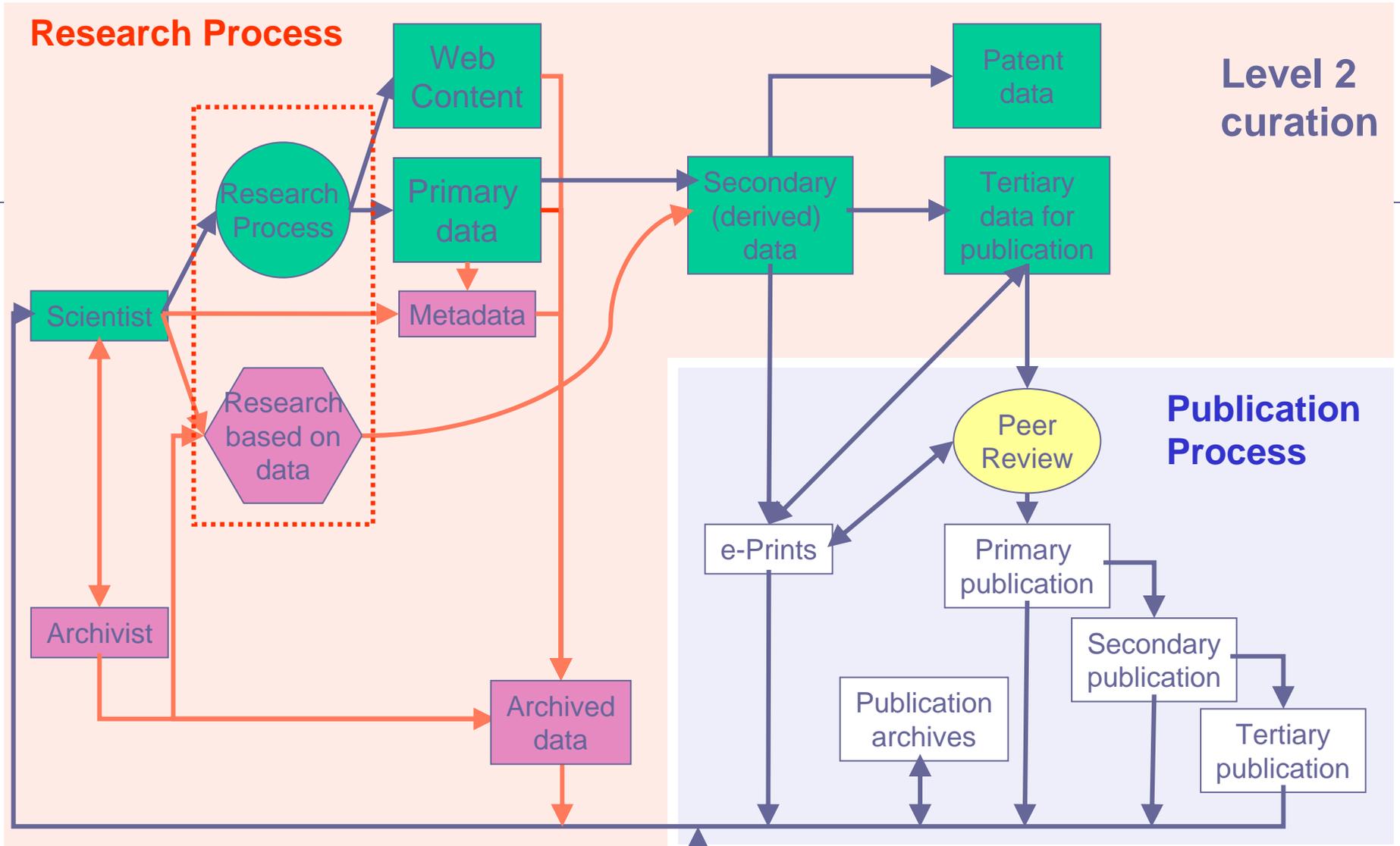
Phase One Progress, March -

- weekly AccessGrid/telecon; two face2face meetings
 - defining programme of deliverables; re-deploying & recruiting staff; planning appointment of full time director in time for Launch
- early ‘deliverables’:
 - **www.dcc.ac.uk** with links, presentations & progress updates
 - **digitalcuration@ed.ac.uk** for contacts & offers of collaboration
- project plan submitted to JISC, late May 2004
- defining R & D programme & services for delivery
 - eg curation architecture; repository of tools & technical information
- engaging curators in existing community of practice

OAIS reference model



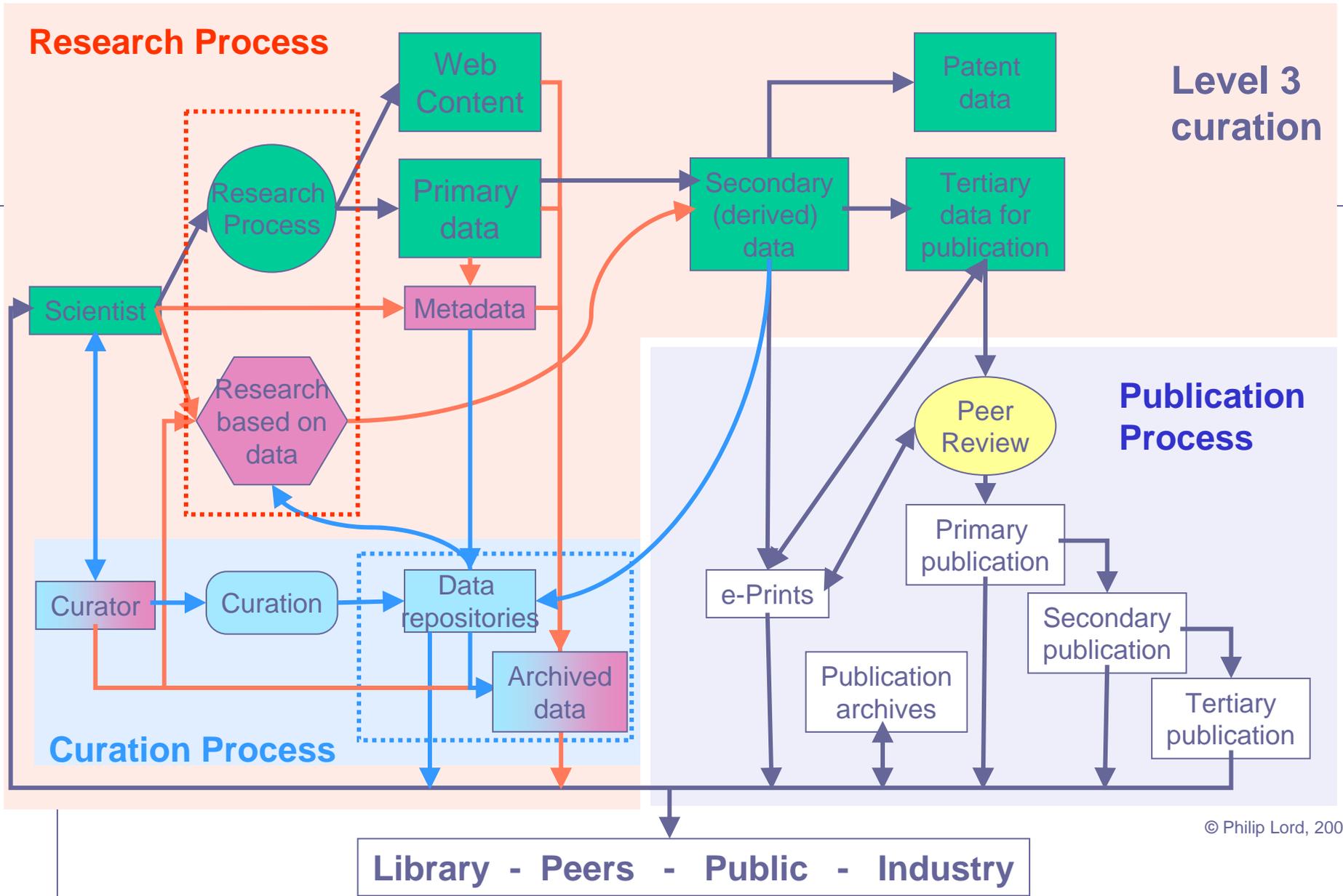




© Philip Lord, 2003

Library - Peers - Public - Industry

Digital | Curation | Centre



Faith in the medium



Faith in the technology

