



Collaboration to Clarify the Costs of Curation



Advanced Electronic Records Institute

24 October, 2014

Cost Models and Sustainability

Neil Grindley – Jisc & 4C Project Coordinator

<http://www.4cproject.eu/>

Project Summary

The Collaboration to Clarify the Costs of Curation (4C) project will help organisations across Europe (and beyond) to more effectively invest in digital curation and preservation.

Vision

The 4C vision is to create a better understanding of digital curation costs through collaboration.

Mission

Our mission is to provide useful, useable resources which support the process of cost management in digital curation.

A quick word about terminology ...

Digital Curation

Digital Preservation

Digital Archiving

Digital Stewardship

Permanent Access

Long-term Information Management

[Digital]... Repositories

Archives

Libraries

Managed Information Environments



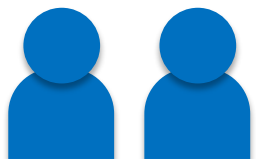
Engagement

Tasks

- Engage stakeholders
- Raise awareness
- Organise meetings
- Promote Research & Innovation
- Build community network



Collaboration to Clarify the Costs of Curation



Assessment

Tasks

- Assess cost models & strategies
- Examine good practice
- Analyse requirements
- Integrate components
- Produce guidance & briefing materials
- Setup costs exchange



Networking & Coordination



Affiliate Partners & Stakeholders



Project Coordination

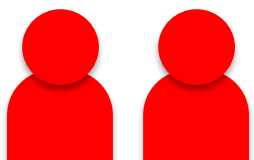
Tasks

- Project meetings
- Project reporting
- EC liaison
- Budget oversight
- Outputs QA

Enhancement

Tasks

- Examine and refine related concepts
- Value
- Risk
- Benefits
- Sustainability
- Economic Reference Model



Outputs



Reports for General Dissemination



Curation Costs Exchange



Reports for European Commission



Submission of Roadmap to the EC



Collaboration to Clarify the Costs of Curation



TIMELINE

May 2011

Nov 2011

Apr 2012

Feb 2013

Preparation



Project Kickoff

Who we are & what we do



iPRES2013

Summer 2013

Contact Stakeholders

Participation, debate, emerging findings



Archiving

Summer 2014

Emerging Resources

Dissemination, legacy, recommendations



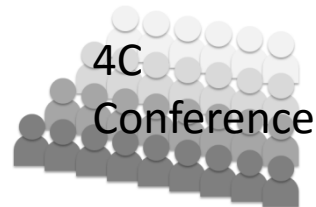
Economics of Digital Curation Roadmap



Jan 2015

Project Close

Oct 2014



Curation Costs Exchange



Why should we concern ourselves about the cost of curation? (What are the stakeholders saying ...)

- Understanding the cost of preservation may mean we can offer realistic and cost effective curation services to others.
- Understanding costs can support strategic planning.
- Understanding costs can support tactical decision-making.
- Understanding costs can provide evidence of cost-effectiveness and value.
- Clarifying and publishing the cost of digital curation can be used to enhance our organisation's credibility. But this must be done along with the context of how the costs were calculated
- Understanding economic drivers can help to strategically align an organisation



4cproject.eu/community-resources





Collaboration to Clarify the Costs of Curation

Go

Login

Register

Site Map

SHARE    ...

Home

About 4C

Work Packages

Community Resources

News and Comment



Community Resources

In This Section

[OUTPUTS AND DELIVERABLES](#)

[FOCUS GROUPS AND
WORKSHOPS](#)

[STAKEHOLDER PARTICIPATION](#)

[RELATED PROJECTS](#)

[GLOSSARY](#)

Community Resources

There is a sizeable canon of research into cost modelling for digital curation. This research has tended to emphasize the cost and complexity of digital curation and preservation, but the research is in many ways preliminary and there has been little uptake of the tools and methods that have been developed—for example, tools to manage and estimate costs have not been integrated into other digital curation processes or tools. The question is why? That's where the 4C project comes in.

The 4C project's mission is to provide useful, useable resources which support the process of cost management in digital curation. Major outputs from the project will include:

www.4Cproject.eu

What we don't want to do ...!

ID	Name	Acronym	Owner
1	Test bed Cost Model for Digital Preservation	T-CMDP	National Archives of the Netherlands
2	NASA Cost Estimation Tool	NASA-CET	National Aeronautics & Space Administration
3	LIFE ³ Costing Model	LIFE3	University College London and The British Library
4	Keeping Research Data Safe	KRDS	Charles Beagrie Limited
5	Cost Model for Digital Archiving	CMDA	Data Archiving and Networked Services (DANS)
6	Cost Model for Digital Preservation	CMDP	Danish National Archives and The Royal Library, DK
7	DP4lib Cost Model	DP4lib	German National Library
8	PrestoPRIME Cost Model for Digital Storage	PP-CMDS	The PrestoPRIME project
9	Total Cost of Preservation	CDL-TCP	California Digital Library
10	Economic Model of Long-Term Storage	EMLTS	Rosenthal, D.

Come up with another Cost Model ...!

Only 15% of people in the 4C stakeholder consultation indicated that they had tried to use a cost model

What can be done to ensure that models are developed in line with users' needs?

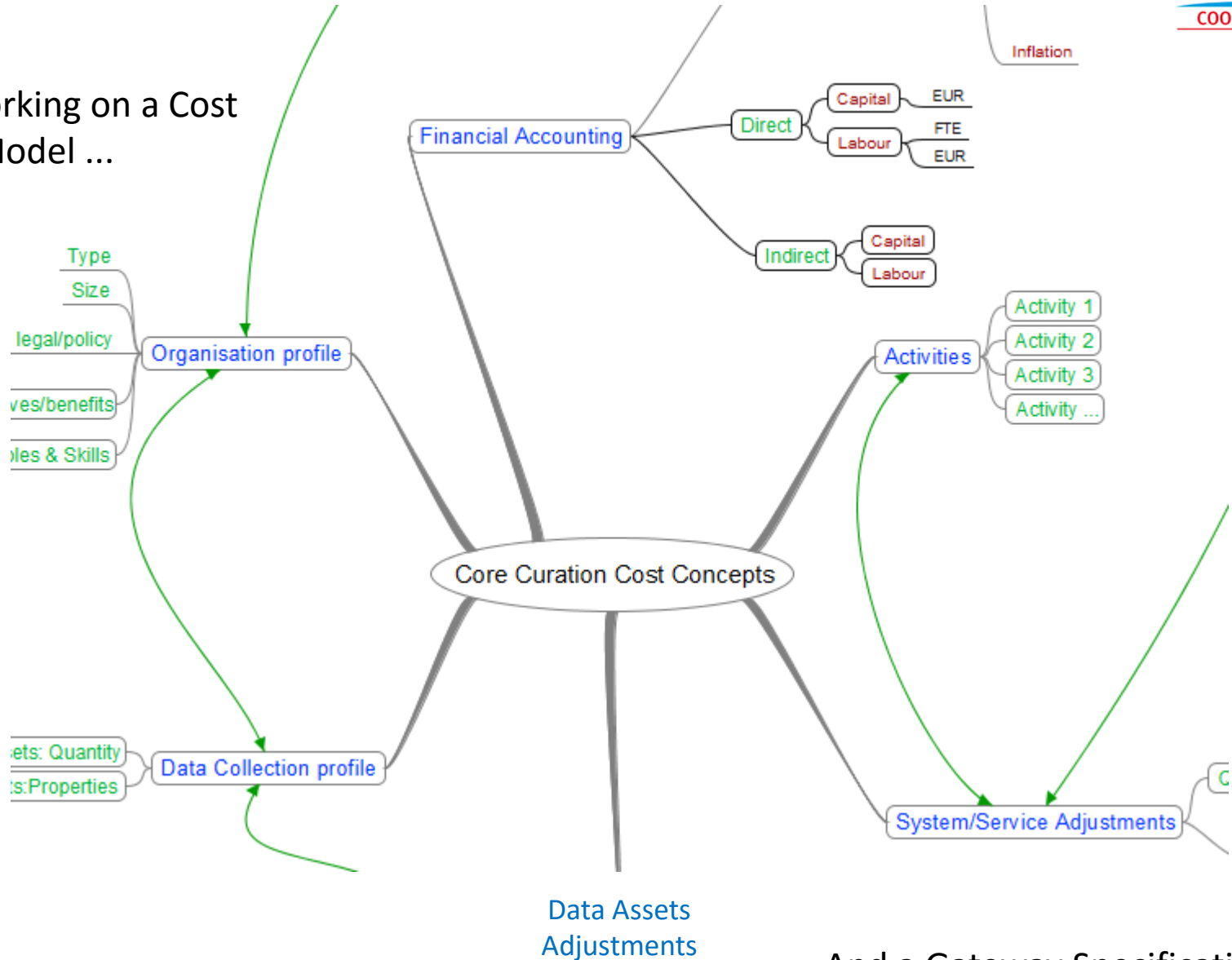
15 drivers for development are listed in the Needs & Gaps report along with 11 recommendations

http://4cproject.eu/component/docman/doc_download/34-d3-1-evaluation-of-cost-models-and-needs-gaps-analysis?Itemid=

Draft good practice proposals for cost model developers

1. Use a standardised definition of digital curation
2. Limit the purpose of the model and define clearly the expected users of the model and its scaling capacity
3. Start out and continue to prioritise simplicity; be explicit about limitations on accuracy
4. Limit the time scope
5. Use simple formulae
6. Implement the model in a simple and widespread tool

We are working on a Cost Concept Model ...



So the Cost Concept Model and Gateway Specification should help in various ways ...

- It will be a one-stop shop for thinking about the components of cost models
- It will enable comparisons of various existing cost models
- It will help people to design their own modelling approaches
- It will help to build consensus around definitions and terminology
- It will be an accommodating structure to showcase and raise awareness of future cost modelling work
- It will feed into the 4C Roadmap work which will recommend future activity in relation to further clarifying the economics of curation

Other Concepts and models we are developing ...

A Taxonomy of Indirect Economic Determinants

authenticity

benefit

efficiency

impact

innovation

interoperability

quality

reputation

risk

sensitivity

skills

sustainability

transparency

trustworthiness

value

Other Concepts and models we are developing ...

A Taxonomy of Indirect Economic Determinants

authenticity

benefit

efficiency

impact

innovation

interoperability

quality

reputation

risk

sensitivity

skills

sustainability

transparency

trustworthiness

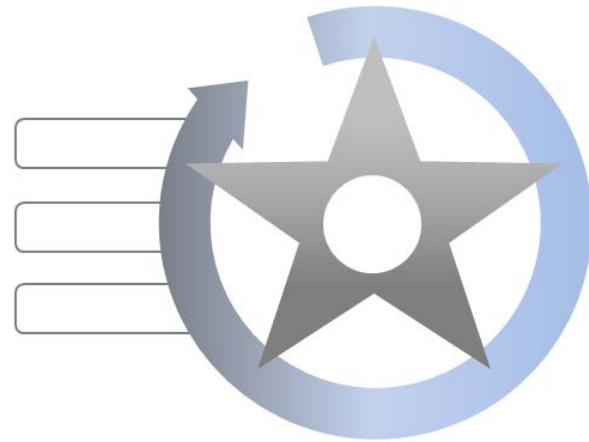
value

There will be various 4C outputs including reports and a Roadmap but let's focus on two in particular ...



The Curation Costs Exchange (CCEX)

The Economic
Sustainability
Reference Model
(ESRM)





We need to think carefully about what exactly *is* the problem that we are trying to solve ...

If they want to, organisations can work out how much it costs them to manage their digital assets

4C Data Gathering Exercise
Organisation A

- Curation Categories
- Pre-Ingest
 - Ingest
 - Preservation Planning
 - Data Management
 - Archival Storage
 - Access
 - Administration

Accounting Principles

Labour Direct	Labour Indirect	Capital Direct	Capital Indirect	Note
---------------	-----------------	----------------	------------------	------

0	0.2	0.1	0.0	0.7	0.6	1.3			€ 5,000	€ 5,000
0	0.2	0.1	0.0	0.7	0.6	1.3			€ 5,000	€ 5,000
0	0	0.1	0.0	0.0	0.3	0.3			€ 1,000	€ 1,000
0	0	0.1	0.0	0.0	0.3	0.3			€ 1,000	€ 1,000

Time Period
2012

Total Cost
€252,000

4C Data Gathering Exercise
Organisation B

- Curation Categories
- Ingest
 - Curation
 - Access

Cost Categories					
Hardware	Software	Employment	Accommodation	External Services	Transfer

Accounting Principles			
Direct	Costs of Service	Absorbed Indirect Costs of Service	Unabsorbed Indirect Costs of Service

Time Period
2012

Total Cost
€645,683.26

4C Data Gathering Exercise
 Organisation C

- Curation Categories
- Ingest
 - Data Management
 - Archival Storage
 - Preservation Planning
 - Access
 - Administration
 - Common Services

Accounting Principles

<i>Labour Costs</i>	<i>Capital Costs</i>	<i>Offset By Revenue</i>
---------------------	----------------------	--------------------------

Size of Collection
 393 TB

Time Period
 2012

Total Cost
€15,800,000

4C Data Gathering Exercise
Organisation D

Curation Categories

- Ingest
- Archival Storage
- Metadata Management
- Access
- Administration

Cost Categories
Product/Service

<i>Requirements</i>	<i>Customisation</i>	<i>Integration</i>	<i>Installation</i>	<i>Training</i>	<i>Support</i>
---------------------	----------------------	--------------------	---------------------	-----------------	----------------

Cost Categories
Hardware

Time Period
?

Total Cost
€349,665

4C Data Gathering Exercise
Organisation E

Curation Categories
Digital Archiving

Cost Categories			
Hardware	Software Maintenance	Software Development	Staff Other

Time Period
2007

Total Cost
123,000

Time Period
2012

Total Cost
348,500

4C Data Gathering Exercise
Organisation F

Curation Categories
Long term Digital
Preservation

Size of Collection
2 TB

Assumption
€8k per TB
per year for
Storage Costs

Accounting Principles
Development & Improvement
Operation

Staff
Development, Technical Support
Training, Communications, Public Relations
Expenses
Software Design
Software Licenses
Support
External Development
Hardware Purchase
Hardware Operating costs
Graphic Design

Time Period
2007-2012

Total Cost
€205,000

4C Data Gathering Exercise
Organisation G

Curation Categories
Digital Archiving

Content Management
Data Development
Hardware
Production
Systems Development
Delivery
User Support
Overhead
Operations
Management

Time Period
2012-13

Total Cost
€ 3,130,110

So ... What exactly are the problems we need to tackle?

- The random numbers problem - How can we meaningfully compare the numbers that we end up with? [*cost data*]
- Activity based costing versus financial accounting methods
- Describing what the organisation does [*cost metadata*]
- Describing the amount and type of data that is being looked after [*cost metadata*]
- Sensitivity around data – Many organisations are not particularly happy to broadcast what it costs them to manage their data. How can we effectively anonymise the sharing of data?
- Complexity - The detail builds up very quickly across different organisations and it doesn't map together easily
- And we somehow have to make sure that the benefits are presented alongside the costs

What the CCEx can do for you



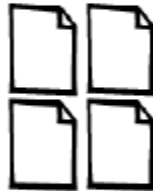
Input your cost data



Compare your expenditure with others



Know how others are spending their budget



Browse through cost models descriptions and comparisons



Know what cost models you should use



Know about other's risks and benefits



Connect with others with the same motivations



Stay informed about all news and events



Know about suppliers and their services

My profile

← → ✕ 🏠

Register

Create an account in the CCEx and become part of the network.

Email

Password

Repeat password

About myself (optional)

Experience with costing ▼

Some experience with costing but not on digital curation costing

Experienced in digital curation costing

Notes on your cost experience

List of top cost activities

←
→
✕
🏠

<http://ccex.org/mycosts/>

Q

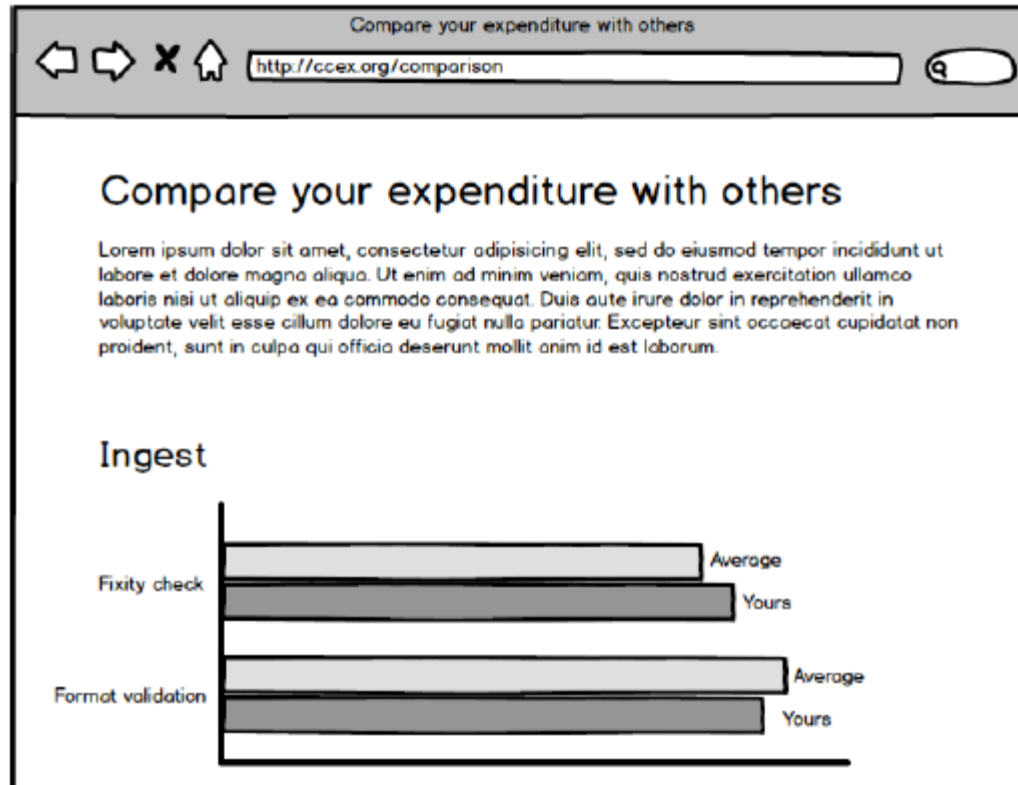
My curation costs

Information about the costs of curation on your organization.

Ingest sub-activities

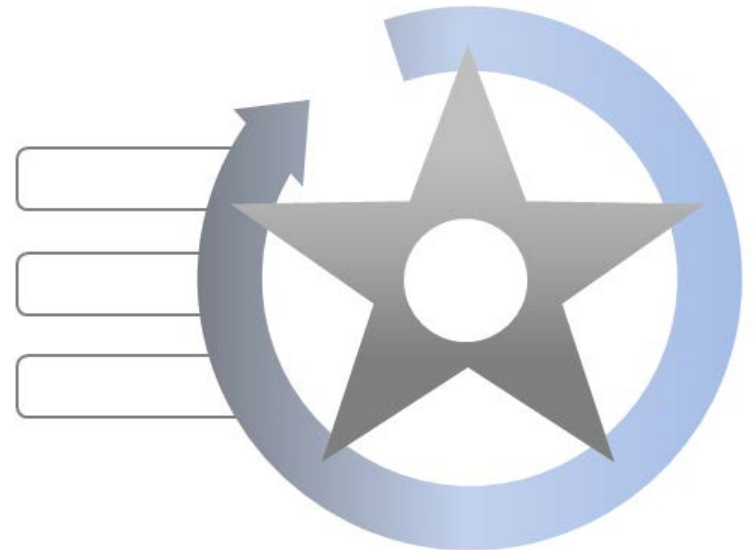
Add
Group by
CSV

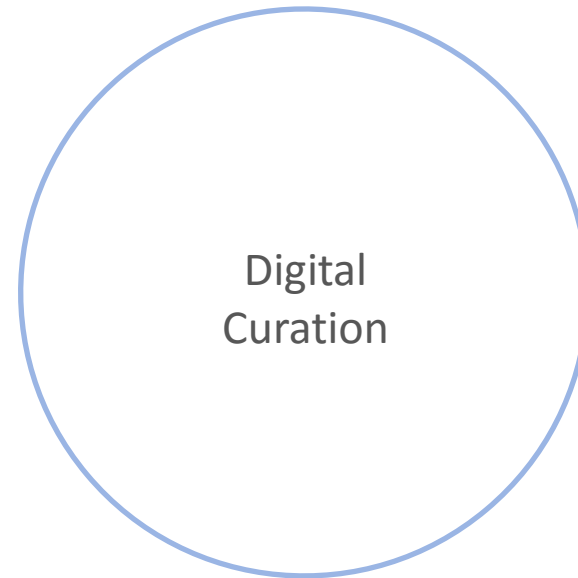
Name	Cost	Breakdown	OAIS entitie	OAIS functi	Raw data
Validation	14100 €	0 sub-activities	ingest	validation	<input checked="" type="checkbox"/>
Incorporation	2200 €	0 sub-activities	ingest	incorporation	<input checked="" type="checkbox"/>
Sum 16300 €					



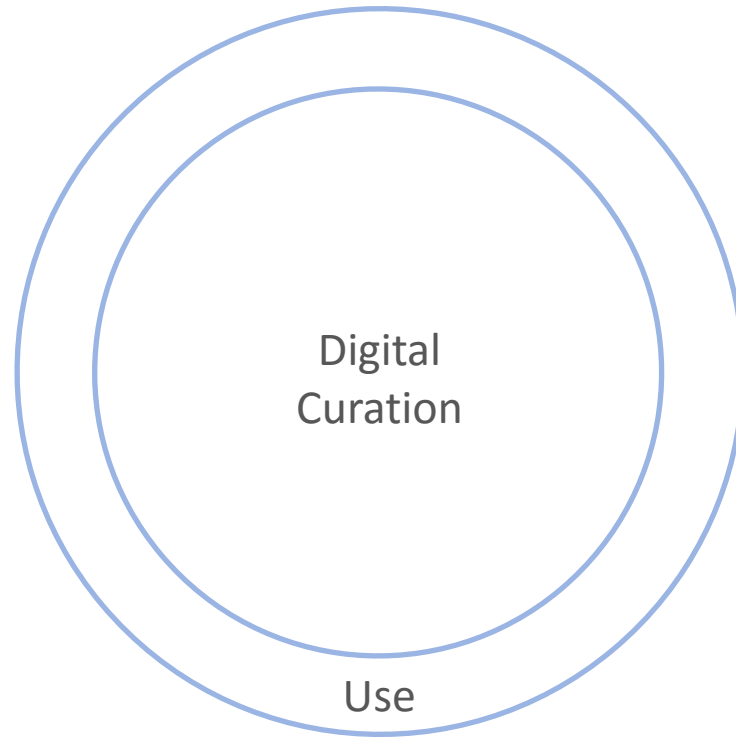
The Economic Sustainability Reference Model proposes that a sustainability strategy requires consideration of four categories of issues:

- The Economic Lifecycle
- Sustainability Conditions
- Key Entities
- Economic Uncertainties

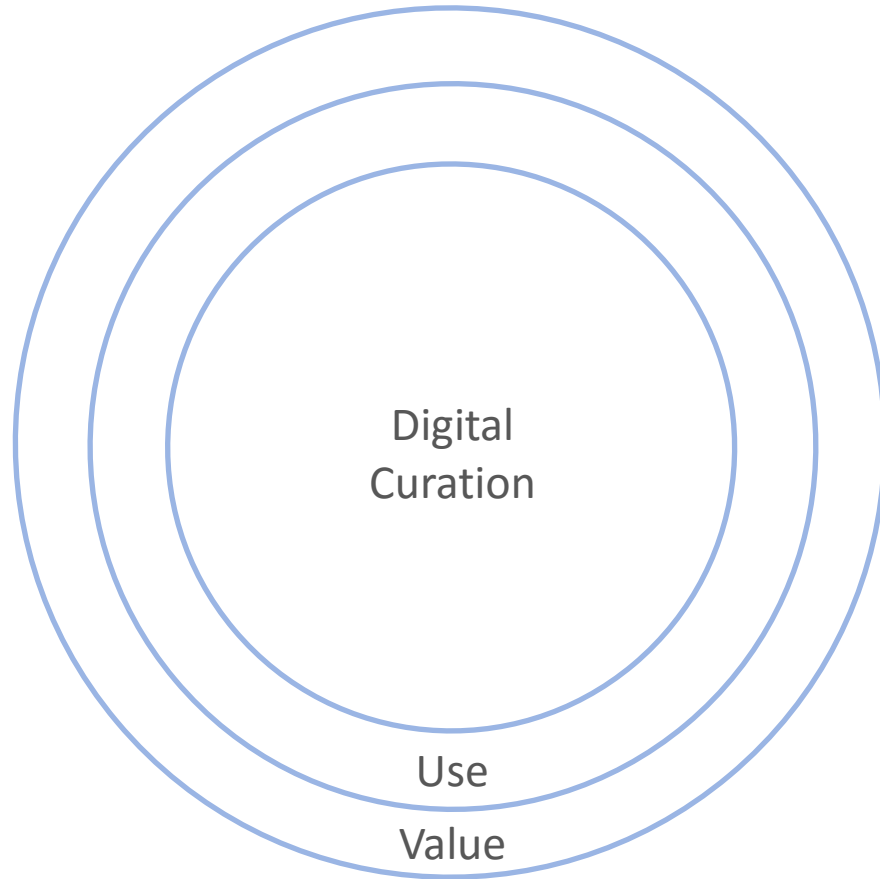




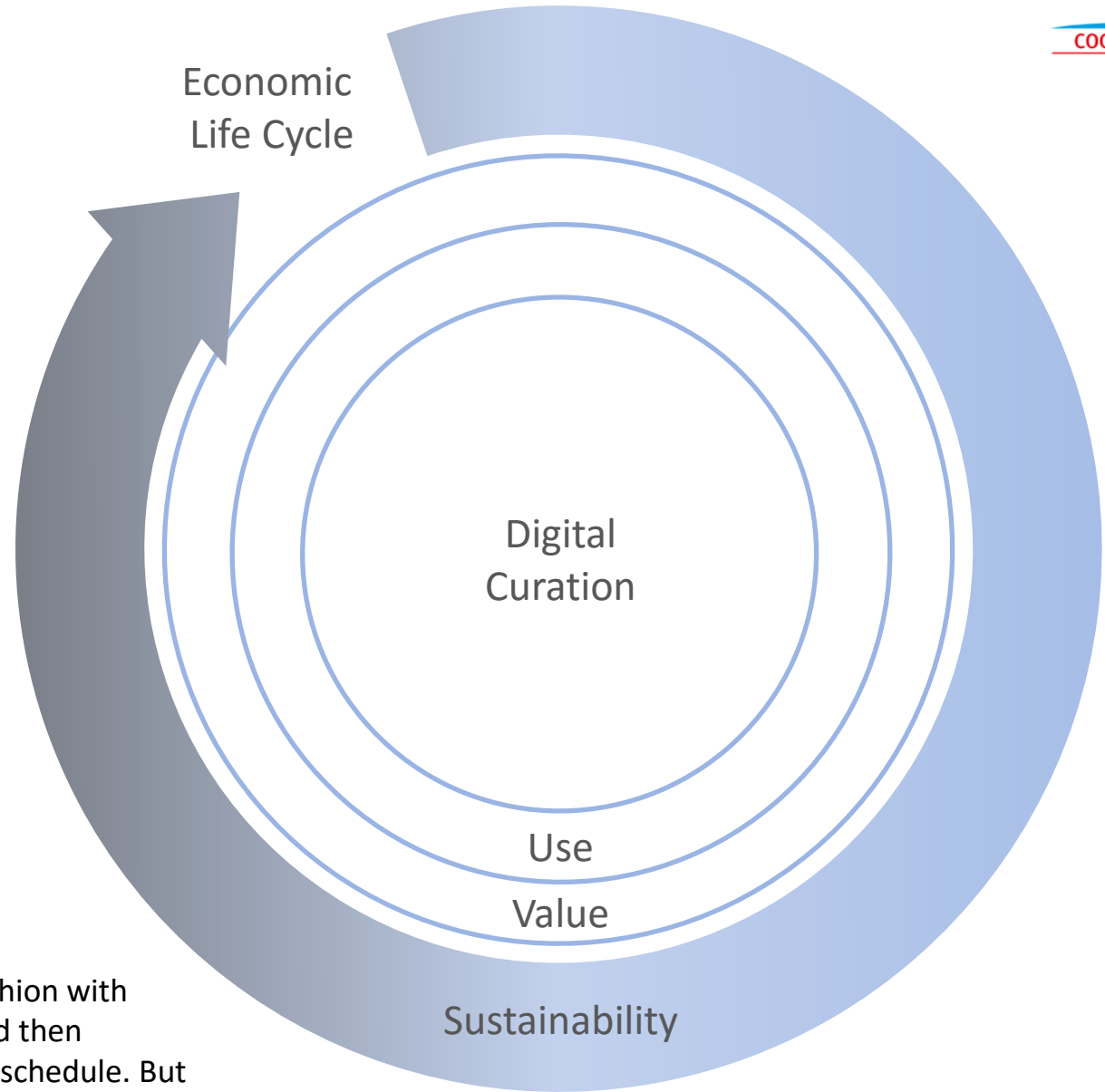
The activity of digital curation is assumed to be the central active component and the engine that will ensure the sustainability of digital assets



Investment into curation will in turn facilitate use (or the potential for use)

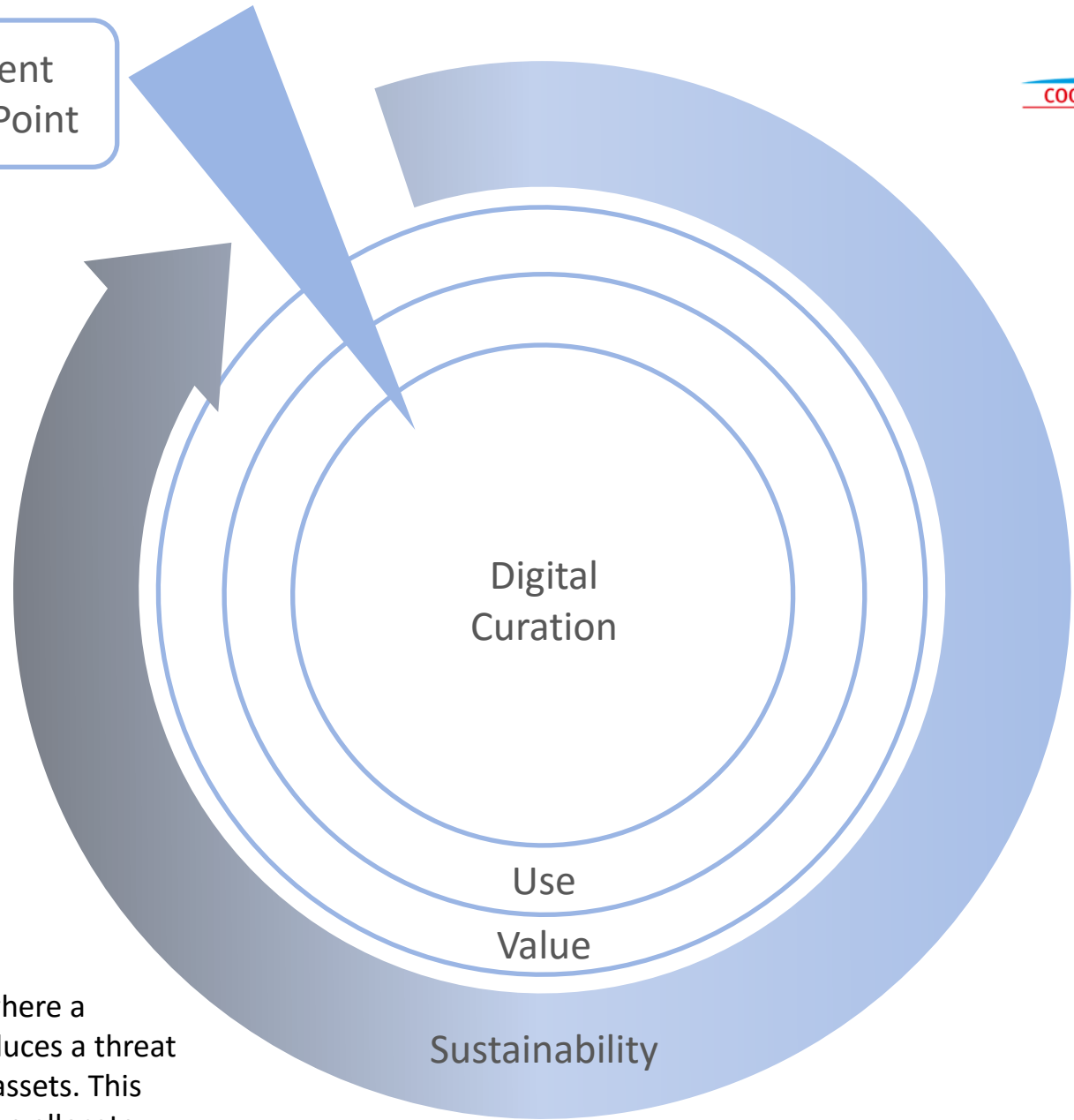


And use (or the potential for use) will realise value, thereby delivering a return on the investment



This could play out in a linear fashion with assets being created, curated and then deleted according to a retention schedule. But in the context of sustainability, it is more likely to be a cyclical process

Investment
Decision Point



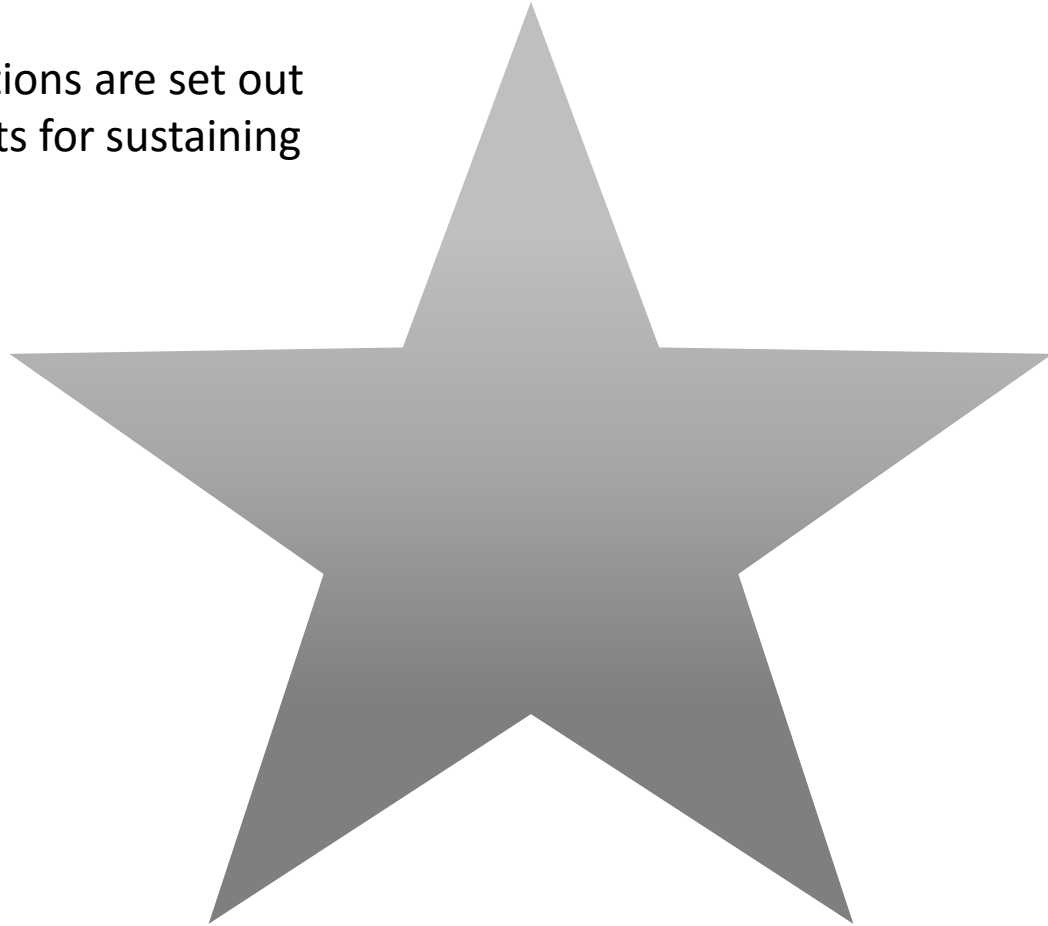
There will be a gap in the cycle where a technical or business issue introduces a threat to the continued viability of the assets. This becomes a decision point ... Do we allocate more resources to tackling the problem?

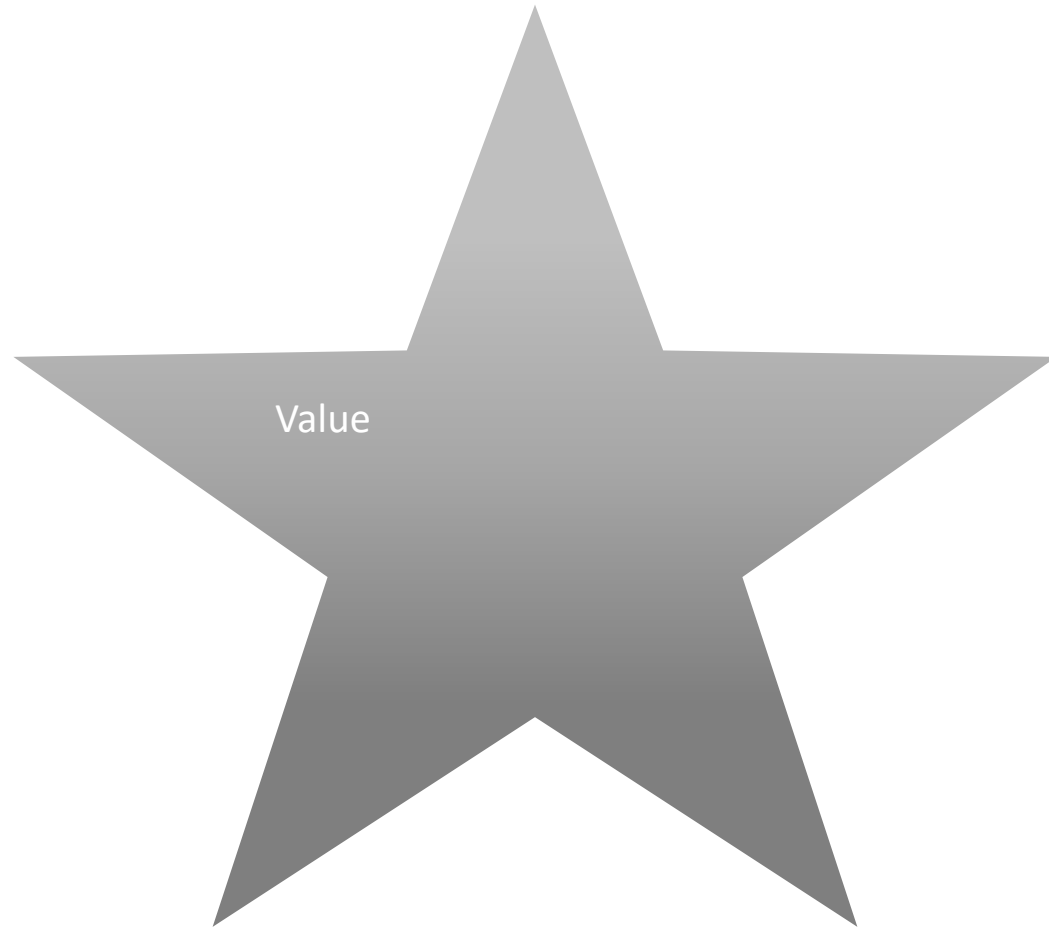


Sustainability Conditions

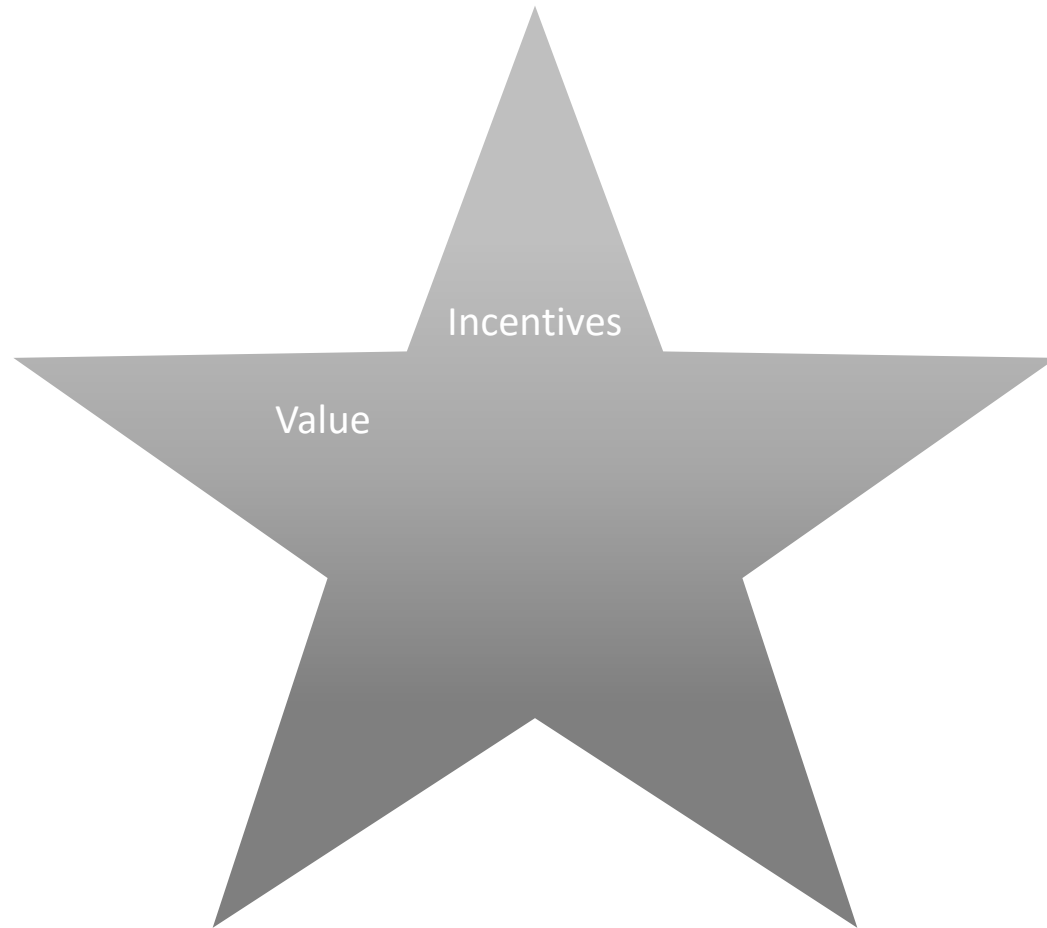


Five Sustainability Conditions are set out to maximise the prospects for sustaining assets

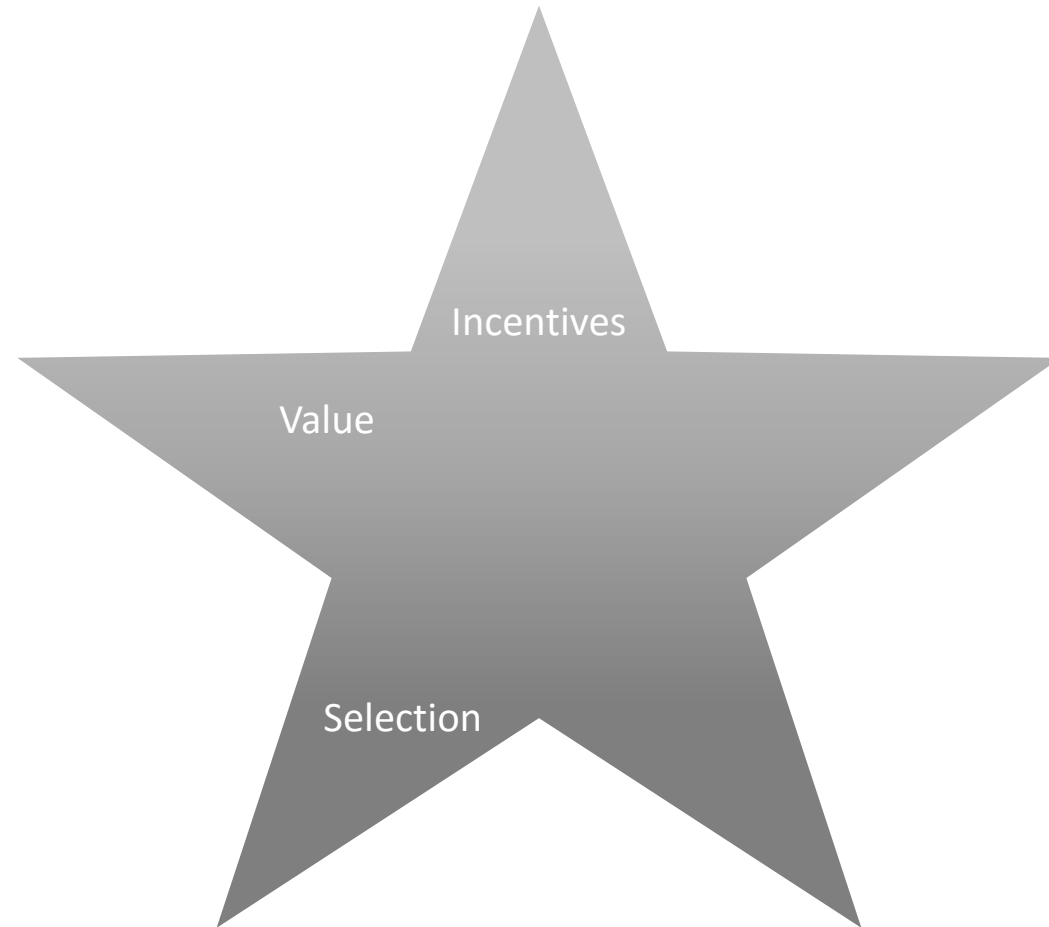




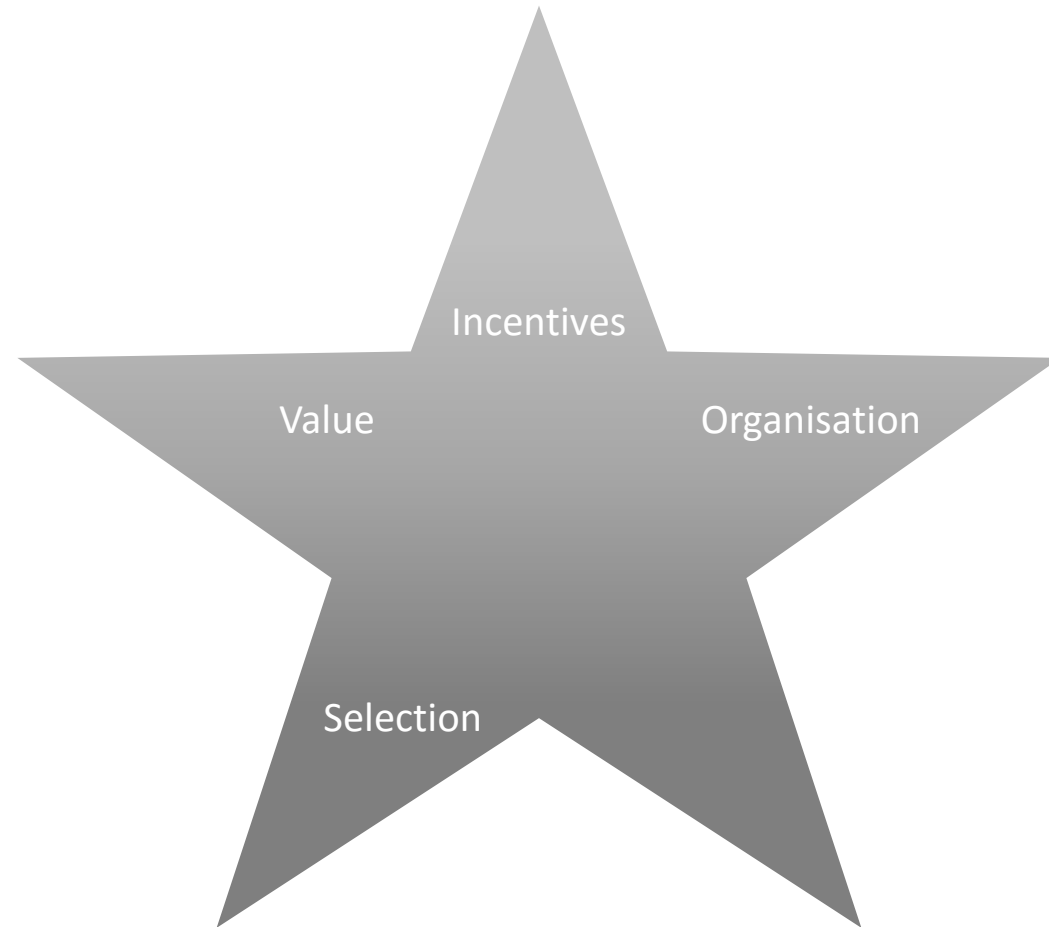
The assets must be understood (or perceived) to have tangible or intangible value



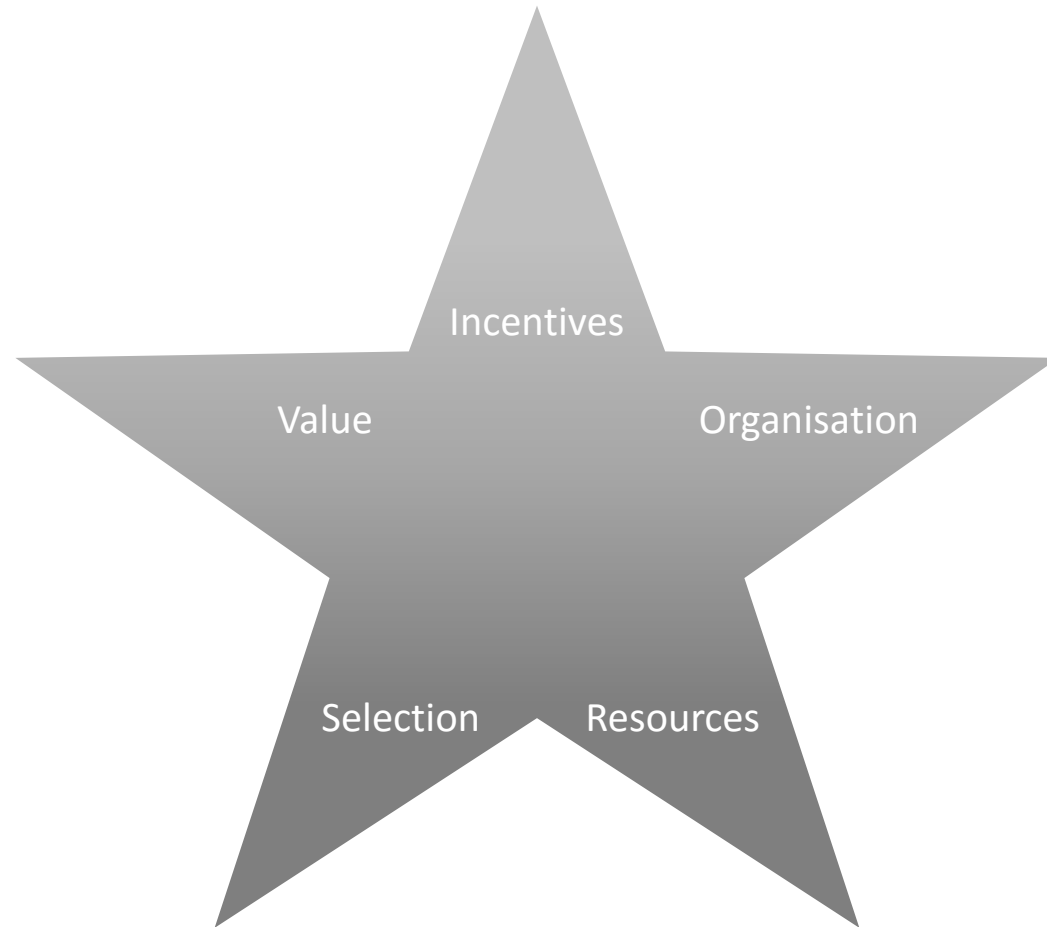
Relevant stakeholders must be sufficiently motivated to support curation



Where resources are scarce then discretion must be used to prioritise curation of the most valuable assets



The organisation should have an appropriate mandate; a supportive governance structure; and be optimally configured to sustain the assets



There must be a sufficient flow of ongoing resources (including financial and human capital) to achieve long-term goals



Key Entities

Three Key Entities are set out which are found in all digital curation contexts. Sustainability requires the nature of these entities to be understood



Three Key Entities are set out which are found in all digital curation contexts. Sustainability requires the nature of these entities to be understood

ASSETS

Every type of digital asset exhibits various attributes or properties that to a greater or lesser extent may affect the how they are curated

STAKEHOLDERS

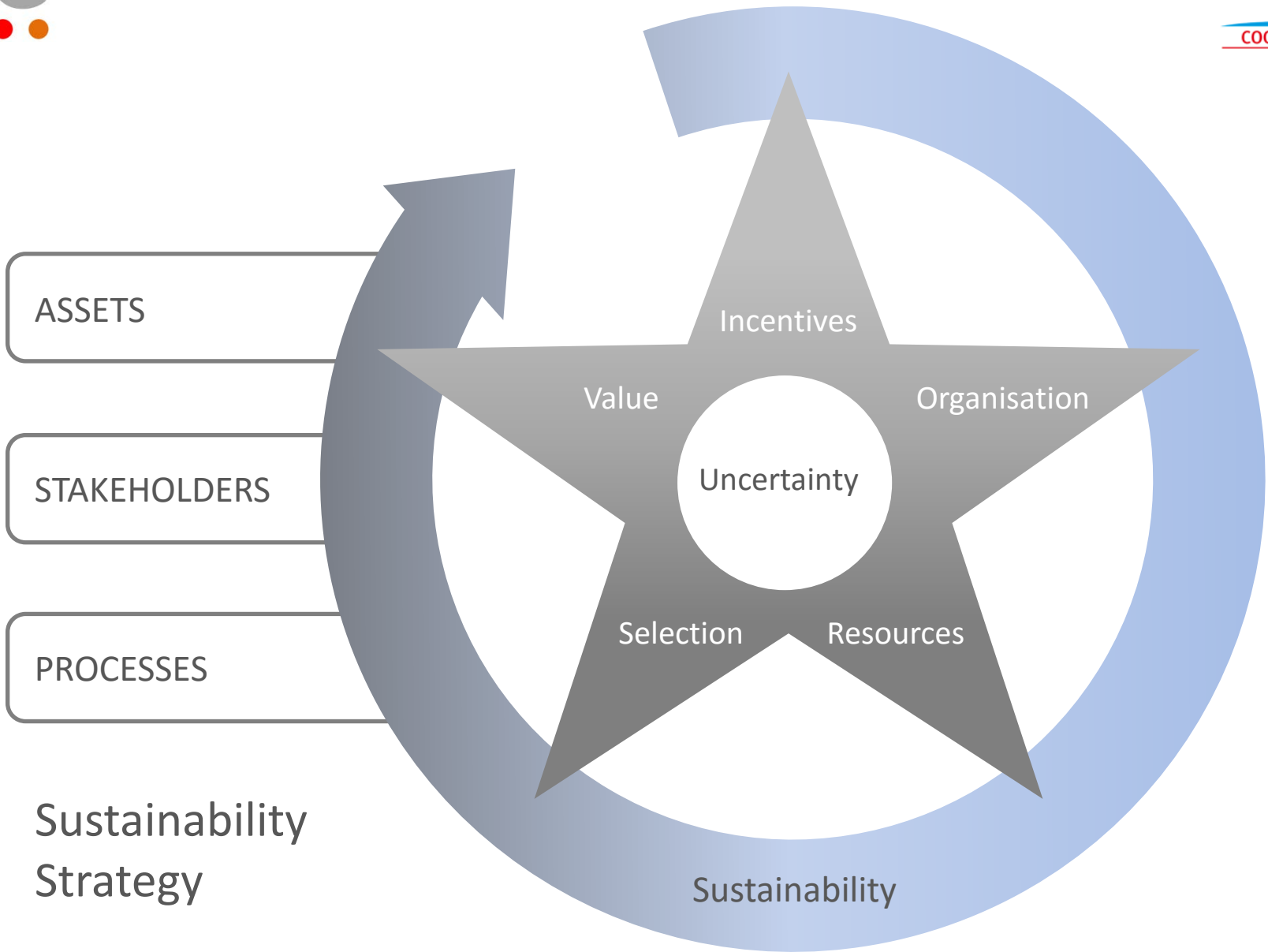
The stakeholder ecosystem for digital assets can be complex and the supply side and demand side should be understood in relation to who is undertaking the curation for the benefit of whom

PROCESSES

The processes involved must be capable of (and optimised for) efficiently enhancing the value of the assets

The inclusion of Economic Uncertainties is an acknowledgement that even the best sustainability strategy cannot accurately predict the future and that some expectation or mitigation of uncertainty (both threats and opportunities) should be built into the strategy where possible





I'll stop there ...

Thanks for listening

Any questions?

n.grindley@jisc.ac.uk