

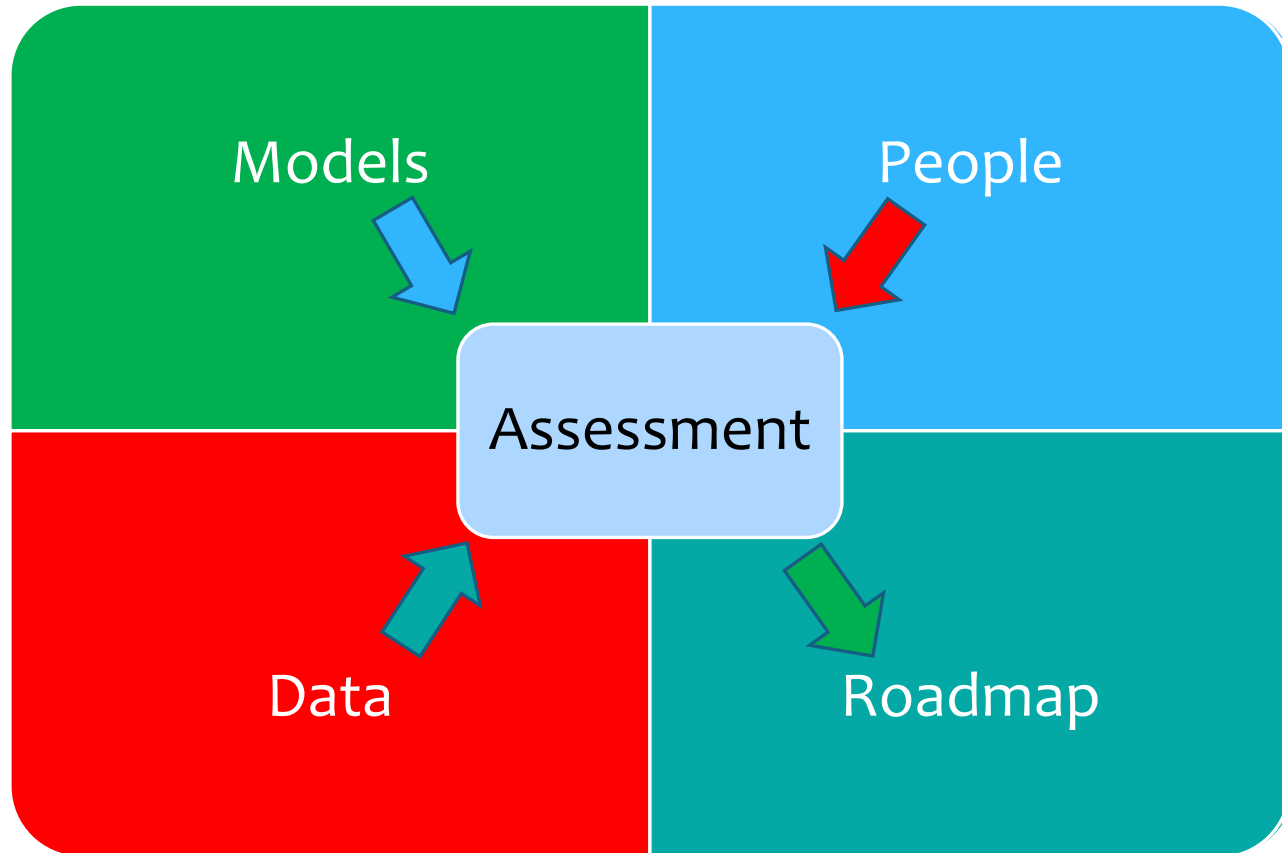
Collaboration to Clarify the Costs of Curation

Initial results from 4C's work on assessing current models and the consultation of stakeholders on digital curation costs

4C WP3 T3.2

KBDK, DNA, HATII-DCC, SBA, UEDIN-DCC, KNAW-DANS, NLE

Collaboration to Clarify the Costs of Curation



Collaboration to Clarify the Costs of Curation

Evaluate existing cost models and produce a needs and gap analysis report

- What's already out there?
- What do the stakeholders want?



Description of how to bridge the gaps



Collaboration to Clarify the Costs of Curation

Develop a cost concept model and gateway requirement specification

Purpose?

→ Describing an area in a conceptual way

Product?

→ A requirements specification?

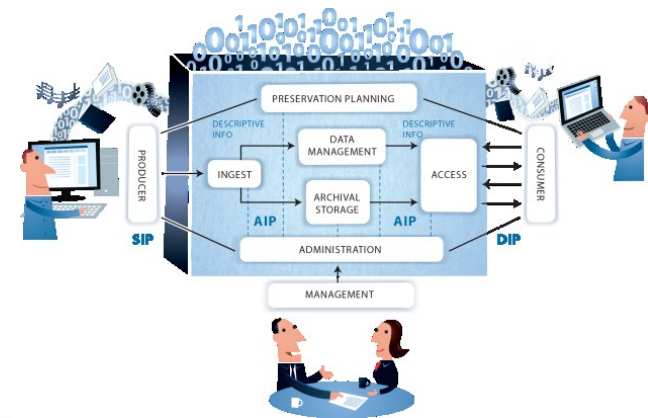
→ A checklist?

→ OAIS for costs of curation?



Utility?

→ Input for the Roadmap, specification for the development of cost models, reference description for the costs of curation



Collaboration to Clarify the Costs of Curation

Develop a submission template and specification for the Curation Costs Exchange (CCEx)

Purpose?

- Gathering cost data
- Understanding cost data
- Allowing people to compare their costs



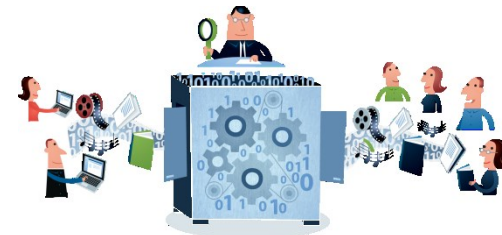
Product?

- An online machine



Utility?

- Sustained networking, service; sandpit



Work Package 3 – Assessment – is about:

→ understanding the needs of stakeholders

→ understanding cost data

→ creating a way of handling cost data in a standardized and useful manner

GAP

- Evaluate existing cost models and produce a needs and gap analysis report

Model

- Develop a cost concept model and gateway requirement specification

Machine

- Develop a submission template and specification for the Curation Costs Exchange (CCEX)



Back to:

Initial results from 4C's work on
assessing current models and the
consultation of stakeholders on
digital curation costs

Contents

1. Theoretical assessment of existing cost models
 1. General description and common traits of the models
 2. Strengths
 3. Weaknesses
2. Results from web consultation of stakeholders
3. Future work

List of evaluated models

ID	Name and acronym	Owner	Type
1	Cost Model for Digital Preservation	National Archives of the Netherlands	Cost
2	The NASA Cost Estimation Tool (CET)	National Aeronautics & Space Administration	Cost
3	LIFE ³ Costing Model	University College London; British Library	Cost
4	Keeping Research Data Safe (KRDS)	Charles Beagrie Limited	Cost, Benefit
5	Cost Model for Digital Archiving	Data Archiving and Networked Services (DANS)	Cost, Benefit
6	Cost Model for Digital Preservation (CMDP)	Danish National Archive; Royal Library, DK	Cost
7	DP4lib Cost Model	German National Library	Cost
8	PrestoPRIME Cost Model for Digital Storage	PrestoPRIME project	Cost
9	Total Cost of Preservation (TPC)	Digital Library of California	Cost
10	Economic Model of Long-Term Storage	Rosenthal, D	Cost
11	Economic Sustainability Reference Model (ESRM)	Rubridge, C; Lavoie, B.	Economic

General description and common traits of models

They are almost all

- * Cost models
- * Using an activity based approach and include an activity checklist
- * Including all types of costs – indirect costs, direct costs, they differentiate labour costs, calculate inflation/deflation, depreciation/amortization
- * Break down activities in a standardized way (OAIS)
- * Very detailed!
- * Limited take-up...

Weaknesses of the models

Limited take-up due to?

- * Not lack of easy to use, quick start guides. Many of the tools reviewed had very detailed user guidance - some guides coming in at 90+ pages.
- * Lack of concise, high-level overviews using plain language and common descriptive elements detailing what the tools can provide
- * Lack of modularity – cost models generally lack the possibility of omitting certain activities without this having an impact on other activities
- * Lack of customization to local set-ups
- * Models are stand-alone tools that don't allow automatic feeding from other systems of an institution.

Weaknesses of the models

- * They're not all finished/covering all functional entities
- * The use of variables is inconclusive – some models deal with complex formats, migration, accession, quality of repository, etc., while others don't - inconsistency
- * Lack of accuracy – The NASA CET model, which may be one of the most accurate ones, implies 25% of inaccuracy and estimates that this is about the best that can be done.

Strengths of the models

- * Include all types of costs
- * Use standards (accountability principles, OAIS)
- * Detailed
- * Documented

Conclusion from our reviewers: “Overall, the quality of the models we reviewed is very high and there are a number of excellent features in each that will help specific user communities get a good grasp on their curation costs”.

What do the stakeholders want?

-Web consultation

- * Goal
 - * Examine practices for assessing costs and benefits of digital curation and obtain information of stakeholders' needs in this area
- * Consultation
 - * 296 invitations sent out to various stakeholder groups from the private and public sector
 - * 1. set of questions to collect general information about the nature of the organisation - 76 responses (25%)
 - * 2. set included optional questions to collect more specific information about incentives for curation, nature of information assets and digital curation activities, accounting and budgeting - 46 responses (16%)

What do the stakeholders want? -Who are they?

- * 37% Memory institution or content holder
- * 20% Commerce (DP vendors, publishers, SME, Data intensive industry)
- * 13% University
- * 9% Government agency
- * 9% Big data science
- * 12% Other
- * 70% indicate that public funding is the main funding source for curation activities
- * 61% indicate that digital curation is their core activity

What do the stakeholders want? -current practices

- * Organisations need to account and make budgets for various
 - * types of assets and use cases
 - * timescales for which digital assets need to be curated, from short, over medium, to long term
 - * numbers of files and data volume
 - * activities, in various qualities
- * 2 out of 3 organisations do not break down digital curation costs

What do the stakeholders want?

-cost modelling experiences

- * Most organisations determine costs based on experience
- * 20% use or have tried to use a cost model
- * Main drivers for using a cost model
 - * Budgeting, comparison of scenarios, assessment of benefits/value → efficiency, money-saving
- * Stated challenges with cost models
 - * Inaccurate, difficult to adapt, don't map well to activities in organisation, lack clear definitions of activities, difficult to use, missing guidance

What do the stakeholders want? -stakeholders' recommendations

- * How can cost models be improved?
 - * Better definition of activities, more detailed breakdown of activities, distinguish between cost types e.g. fixed and variable costs, model complex objects better, refine cost drivers, develop software to facilitate modelling, improve reliability and accuracy, adaptability to local scenarios...

“By being relevant to the practical context we operate in, not a seemingly theoretical imagination of what is actually done.”

What do the stakeholders want?

-summary of needs of stakeholders

- * Model various types and amounts of assets and use cases (flexible and scalable)
- * Sound definition and breakdown of activities and description of the quality of the activities
- * Sound definition and breakdown of cost types and well defined accounting principles
- * Account for past/current costs; estimate future costs
- * Well documented and easy to use
- * Assessment of benefits/value

NEXT STEPS

- * Analysis of the stakeholders' needs
- * Analysis of the gaps between stakeholders' needs for cost modelling and the capabilities of the models
 - * Product: Needs and gaps analysis report
- * Input to:
 - * Concept cost model
 - * CCEx machine (deus ex...)
 - * Roadmap – guidance on best practices in cost modelling

Resources

and questions????

Please go check out <http://4cproject.eu/> that has blogs, useful information, milestones, deliverables and other interesting community resources!

Web consultation report:

<http://4cproject.eu/community-resources/output-deliverables/d2-1-baseline-study-of-stakeholder-stakeholder-initiatives>

All the funny men and women used as illustrations in this presentation are published under Creative Commons Attribution 2.5 Denmark, cf. www.digitalbevaring.dk – obrigado!

