ENUMERATE Focus Group Meeting on Digital Preservation

Monitoring and Measuring Digital Preservation practices in the EU cultural heritage domain

Place and Date
17 January 2013, The Hague
Venue
Digitaal Erfgoed Nederland (DEN)

Participants:

Name	Organisation
Andreas Rauber (AR)	Technical University Vienna: 10/15 year in Digital Preservation research (only afternoon session)
Barbara Sierman (BS)	KB, The Netherlands: digital preservation manager; R&D department; small advisory team; 2 EU projects; auditing certificate organisation
Gerhard Jan Nauta (GJN)	DEN: working on ENUMERATE; thematic survey and methodology
Henk Vanstappen (HV)	PACKED: Flemish expert centre for digital heritage
Marco de Niet (MdN)	director DEN: Dutch expert centre for digital heritage
Marius Snijders (MS)	PrestoCentre, a membership organisation; founded by 5 European broadcasting orgs; data gathering; EU project Presto4U focused on institutions with smaller collections
Maureen Pennock (MP)	The British Library: digital preservation manager; IT oriented; working on new corporate/organisational strategy; interests: focus on strategy
Mette van Essen (MvE)	National Archive, The Netherlands: digital preservation researcher; framework services for National Archive, The Netherlands; digital preservation is one of the services
Natasha Stroeker (NS)	Panteia: contractor in the ENUMERATE project
Neil Grindley (NG)	JISC, a funding organisation; national data organisations; partnerships with f.i. universities; 4C project
Robert Gillisse (RG)	DEN: Senior officer, digital preservation; working together with Henk Vanstappen on scoremodel for smaller and middle-sized organisations, for auditing their digital preservation activities

Introduction (10.00 -10.30)

The focus group session starts with a presentation by MdN about the goals of ENUMERATE, the results of the Core Survey and the set-up of the Thematic Survey.

The focus of today lies in:

- improving the understanding of digital preservation practices and policies
- extending the statistical framework of ENUMERATE with indicators for digital preservation

Part I: The need for quantitative data in the domain of digital preservation and digital heritage collections / Existing tools and methods

10.30-12.00: Discussion on previous work and the usefulness of monitoring digital preservation (DP) policies and practice

- ***What are the main motivations for memory institutions and other stakeholders (e.g. policy makers, funding agencies) to monitor the status of digital preservation?***

MdN: What is the importance of metrics about DP in your institutions?

MP: In the British Library the use of metrics is experimental. Previous approach more loose, there is no ongoing assessment of meeting our objectives. It is essential at this point because we want to reach a cultural change across the organisation. Must monitor that. Otherwise risk is too great not succeeding. This is linked to 2020 vision of corporate strategy. MdN: Management driven or from the department? MP: It was my own initiative. Growing knowledge, strategy, agenda 2020, content strategy: well-constructed, dependencies between the strategies, if it does not succeed you have a gap. Metrics used internally - no annual reporting. Set of strategic priorities, +set of associated actions, metrics link back to these both. Monitor how we operate. Planning is a big part of it. Monitoring progress of the metrics towards the plan, every 6 months it should be done. Metrics will be included on a higher level in the corporate strategy.

MdN: this is what we are interested in (high level types of metrics; something that makes comparison possible in between institutions).

BS: we are an advisory group in the KB; we don't write the strategic part of the DP in the KB. MdN: Would you advise to focus more on metrics for measuring success? Good to have clear goals; it starts with objectives. Objectives for KB: to get digital preservation implemented in the KB, who is doing what on DP. Now in 1 part of the organisation (product support), but it should be more merged with the department of access, supplies etc. DP starts at the front office. If this is not the case, than its too early for metrics. First the goals should be clear, and then the next step is the metrics.

GJN asks MP: is the introduction of metrics in the BL promotional too? Are these metrics also meant to "educate" staff? "Ok, so we've got to focus on this or that…" MP: I expect

there will be an element of that. We will use the information to address the areas where we are not succeeding as much as we'd like to… To get people involved. We are an R&D team. Operations is a different department. We focus on technical guidance and advise. The strategy covers all different aspects, the entire life cycle. Get all of the library on board is one of the priorities/goals. Cultural change. Metrics on how good are we in embedding (DP) best practices across in the total organisation.

MdN: this is in line with what we are trying to achieve in ENUMERATE: an overall understanding of the topic (DP) across the organisation.

BS: We start with the policies. In the annual planning reprort: DP is one of the core streams for the next year. How are we doing… etc. Should be transferred in actions.

MdN: Ok, so the KB wants first to get policies in place. Getting policy more evidence based? What kind of evidence could the KB use to develop the policy? How organised? BS: We need to start; so for now I don't know, we are at the beginning of the process. KB is doing a lot of things that are related to DP, but that is unclear. She wants to make it more clear.

NG: has a couple of concerns related to DP and JISC in the next 5 years or so and metrics might help in getting more clarity. The first one is: … EU and its perception on DP. The last DP call did not have much submissions. What's the reason for this? Is DP not the focus anymore? But we [JISC] feel: There is a great deal of work to do. We need demonstrable facts and figures to show what are the reasons that DP activities are very much needed. A lot of DP strategies are drawn up. We need arguments towards funders!!!

MdN: MS do you recognize this conception of the EU about DP? [MS was working for the European Commission until recently.]

MS: There was a survey on audiovisual material a few years ago [2005, survey 'TAPE'; survey on long-term access to audiovisual collections], status of the audiovisual collections, status of digitisation. The influence of such numbers is really big. The importance of collecting numeric data to influence policy and funding organisations is very large. Is it fair? EU needs numbers to build their policy case, the further in time the less they care about how the numbers are gathered. The EU wants to see the impact of their policies (NUMERIC/ENUMERATE). Policital game play behind: they prefer positive outcomes. The numbers get a different meaning on institutional versus aggregate level. At PrestoCentre: not making a case for funding, but making a case for cost reduction. Collect from large samples: calculating costs on large amounts (migrated audiovisual material). So it is for institutions starting up; they get a better rule of thumb about costs. Different focus than the EU had. Data are biased, start with large organisations. Done the job, not so very much interested in delivering the data. Ask 3 persons in one institution - e.g. Nl. inst. Sound&Vision - for the costs, you get three different answers depending on different definitions. Our questionnaire: 20 questions; whole questionnaire: 25 minutes about. In the end: 5 half-filled in results.

MdN: Neil you said you had several concerns. And you mentioned one…

NG: yes, just 2 others… JISC has gone through a transition as an organisation. Economic constraints: JISC has less money than it used to have. Priorities must be set; where to put focus in 3-4 years. Each of the JISC programs has already gone through a business review process. The result for programs such as mine [DP] is that we need to get practical evidence

on impact. Impact on stakeholders in general. Who the stakeholders are varies from program to program. Deliver benefits to higher education institutions (in partnership with libraries, museums, and archives - principal stakeholders: universities). Impact is a difficult problem, always has been. Metrics give clarity …internally. But also sector wide.

MdN: [something that came up at a UNESCO meeting, which struck me:] we are turning DP too much into a professional issue. It should be a societal issue; and not so much an inner crowd type of thing. In-crowd leads to misconceptions; ideas about the severity of some of the issues we are dealing with. Maybe we don't come across that well if we are explaining our work.

NG: We don't do DP as a bureaucratic exercise. We don't simply do it because our organisations need that as an insurance policy or something like that… We preserve for a purpose, which is to give society opportunities to do things that could not be done otherwise. We should say: preservation is a …human opportunity. Getting there with metrics and following that chain of benefits and then articulating actually, well, is there any kind of metric around…

MP: likes the idea of linking back to society, but it is much easier to manage and to measure your success if you link it more to your business needs and your business drivers. The link to society is a subsequent outcome of your success in meeting your business need. That is what you are able to do.

MvE: [about the National Archive digital repository:] We have all the policies, but we don't have the material yet, because by law they [the government bodies] have to give it after 20 years, and then they will say: o yeah, we should give it to the archive, but then it will be a mess. In the analogue world selection is an issue, in the digital world this is not the case: just buy another hard disk. For us the big thing in DP is being there when the archive is created; it is not the traditional procedure of receiving something and then preserving it. Creation and selection should go hand in hand.

MdN: for (national) archives is DP business related activity or is there also a more philosophical/cultural/societal aspect that is prominent in your thinking of the future?

MvE: should National Archive DP policy be for the NA or for whole government? This is an actual discussion (f.i. on open standards). MP: you have to sell the benefits too, to get them on board. In "their" eyes DP is difficult; frustrating; it costs money… For the NA at this time it is not really an issue: which tools to use for the migration. We are not there yet.

MdN: how can metrics and measuring or monitoring progress help you in your mission?

MvE: The analogue way it is clear what you should do. In the digital way insight in for instance the selection process is more difficult to achieve. Selection for instance: how much does that cost? Benchmarking. MdN: So metrics can be of help in cost reduction (see also MS).

MS: For the national broadcasting archive the narrow business case has not changed that much: we get AV materials and we have to deliver them. But the business case of the people who give us the material and the requirements of society HAVE changed. Different expectations towards DP. TV programmes of 20 years ago you understand that you can't

find them - it will take me two weeks to search…, but programs from a year ago, you must be able to find. Expectations are different.

MdN: Do you report your metrics to society? The collections of the past five years or so, is there any quantitative accountability? To the public? MS: indirectly there is. But it is funny that you need a separately funding programme, and then report within these funding structures, Whereas to the structural funding we get from the ministry it is all implied that we are doing this… There is no actual request for actual data. Nobody asks how many items we lost, for example.

MP: metrics that I would be gathering and checking are not the things that the public is interested in: how many DP workflows have we implemented, that sort of questions. The public is interested in the end results; what is accessible; but not the back of this stuff.

MS: The EU/ministries/the government are interested in: Have you spent your money well? Have you spent it in time? Have you spent all your money? …rather than in the output that the money was meant for in the first place. But there is a change, things are slowly changing.

BS: I think things are changing. If you look at the new ways that projects are funded… under the new pre-commercial procurement program the assumption is that the partners in the project buy or implement the results of the project. In PLANETS we developed tools, but few organisations implemented these tools, which actually is not the intention of such projects. My interpretation: the EU commission tries to persuade participants in projects more to really adopt the products of projects. This would raise quality. Things are changing.

BS: By the way: in my view interest for DP doesn't seem to be less at the EU. EU stimulates us to come up with new ideas.

NG: Some individuals at the EU seem very supportive towards DP. But interest at EU level is going down.

MdN: Policy Support Program CALL EU: DP part of it in an operational way (working plan 2014 EU).

MS: interest for DP in EU is depending on individuals. It is scary. It has been better in the past. The challenge last year (50 million available), only 6 proposals, not many interested institutions. That is why EU is experimenting. What better instruments are there to support and promote DP? DP will not come off the political agenda the coming 10 year. EU is looking / searching for the best instruments etc.

NG: EU may be impatient; there is not enough progress; a lot of things are not implemented. Only doing all R&D projects.

MP: …it leads to a reduction of interest of the institutions. We've done a lot of R&D but now we have to implement it; institutions are looking more at internal procedures.

MdN: This is an important point: can we in some way show that many of the investments in R&D have resulted in something visible, something that can be monitored with metrics?

NG: that is also a JISC point of attention. What money do we put in and what are the results? The R&D that we've done, has it affected the actual practice within institutions? NG mentions

this year's DCC [Digital Curation Centre] Conference where a lot of the papers describe very applied activities. So there was more focus on practices and combining tools.

BS: That is about how to implement DP. If you start from a theoretical/research model, the main goal will be to get at implementation in your organisation, but that costs far more time than to do the research. So it is in a way unrealistic to say: the PLANETS (or whatever) project is finished and how are the results implemented in your organisation, because that [implementation] may take a few years! Results of plans take more time to get implemented. They have to do the implementation next to ongoing business.

MP: Yes, that is a risk indeed. If there are no plans about how to implement R&D results in the institution, things change and go along, you can be too late!

MdN: We've seen that a lot of metrics serve internal purposes in the institutions, but there is a third layer [which is] benchmarking among the institutions. How particular is the implementation of DP in an organisation, or how generic, and what are the consequences for benchmarking. Is benchmarking useful? The National Libraries have been working on this, but how are the prospects of this for instance in archives?

MvE: looking also at things that libraries do. But archives are less object-driven; more process-driven. In a general way: everybody has the same problems, but the implementation is very particular.

MdN: How is that for AV? Was benchmarking one of the motivations to run a survey?

MS: Benchmarking in a very formal structured way, no, I don't know examples of it, but there are a lot of comparisons made between comparable institutions (similar data, amount of data, similar purpose). What is being compared are: workflows; types of equipment; types of vendors; costs (broadcoasting); storage; quality control; migration (DP). And if you talk about digitisation for DP purposes: the initial costs of set up; costs of test to getting at work (which quality needed). Comparison can be at micro-level: what type of quality and what scanner is fit for the job. And then there are comparisons at an international scale.

MdN: Would formal benchmarking be useful?

MS: It depends on how you define benchmarking. Comparison is not needed in terms of: where do I stand as compared to another institution, I think. On a microlevel: especially for vendors, they are very interested to see how their equipment works in a specific institution, or with a specific type of data. Market planners will be very interested.

MdN asks about the score model developed by PACKED and DEN. It is a self-assessment tool and a preparation for auditing. Would comparing yourself as an organisation with others be useful?

HV: 'benchmarking' always makes me think of competing [GM versus Toyota], but that is not how we understand it: comparing with norm standard levels of DP (like OAIS). I see it as risk assessment.

MdN: comparing strategies is that useful in reports on progress? HV: first the question should be answered: do you have a strategy? RG: **Do you have a specific strategy for digital preservation?**

BS: Do you have a DP strategy? But what answer do you expect? One says YES, but actually it is not there. So the question should be answered: what do they mean by a "DP policy/strategy"?

MdN: and should every institution have a DP Strategy? What does it mean if 40% of the institutions say they have one. Is that a meaningful sort of information?

BS: if an organisation has a DP strategy then often there is a separate budget for DP. But then they do not to have be successful in DP.

MP: You need a set of follow up questions: How old is the DP strategy? Who is involved? Is it corporate or routine? Is it updated? Who has made it? It is written? NG: Could you produce a copy of it?

BS: If you ask questions, deep down there is an expectation about what is right and what is not. Is there a standard, when are we successful in DP? There must be some sort of organisation. You need to have thought about backups, storage, file formats etc. But there is no accepted model for a DP strategy, yet.

MdN: But that's where the score model can be useful?

MS: Strategy of regional archive is different from a national library. Here a benchmark can lead to misleading outcomes. Comparing the cost per terabyte for an AV archive with the cost of it in a library does not make much sense. Where institutions are very similar it mighty make some sense. So rather than scaling yourself in between a hundred institutions or so, you should focus on what to you do with it.

NG: The relation between institutions' goals, activities and benchmarking can be rather complicated. NG spoke to someone from the ADP Network, Alabama (LOCKSS): he was basically saying that they did not bother about the precise cost; they decided to focus on what's the most cost effective method of DP. They just got on with it, and did not spend much time in benchmarking and having implemented a strategy. DP was a part of their library agreement. They had a very practical attitude: just go on with it. [Although there is some general knowledge, of course, in the community, about metrics.]

Summary MdN:

- Metrics are probably essential for self-assessment, monitoring, implementation of your own R&D results.
- Then one level up, looking at the surrounding environment of institutions, it is useful for comparison reasons (costs).
- There is some other data that can help you assess yourself: data from others about storage or policy choices, where the maturity model may help in defining e.g. 5 steps. A small institution does not need to have the same ambition as for instance a national institution may have.
- The most problematic part we discussed: accountability to government and society. Perhaps the EU or national government may not value the importance of the topic of DP. It is difficult to find good information to get policy development one step further; to find a legitimation to put more money in DP.

Comments:

BS: The EU Commission/governments may not be interested in DP, because DP in itself should not exist. The discussion should be about: keeping the stuff alive. So if you change your angle of view and say: how much of the digital material is surviving and what are the activities to achieve this …it is probably more convincing.

MdN: "DP is the defining challenge for our generation" (says Nick Poole). We are the first generation to be confronted with enormous amounts of digital material. And if we don't make the proper choices…

MvE: Government thinks it is no problem, because we have a digital archive. Change the questions into: how do we tackle the problem in general.

MS: We should not combine EU commission/institutions and national governments in this discussion. EU bodies are interested in things that cannot be done at national level. DP is part of an innovation or research program; through it the EU might boost the economy. National governments are more interests in budgetary issues [costs] - they spend structural funding on "keeping something" - and they are interested in the relation with risk and loss. That combination: budget, risk, and loss… and the information that they need to be able to create appropriate policies and budgets, is very important for them. We need to be able to provide answers: What do you get for this money? Put it in a certain frame: if you do not invest X number of money, this part of the collection is at risk (a way of presenting the problem).

MdN: is it realistic to think that DP practices will move more from the national level towards a European approach? E.g. the Comité des Sages report said: is it an idea to use Europeana as a dark archive? Currently the whole organisation and funding of DP is far more on a national level. Research is international, but the practices, will that move more towards European activities?

MS: Europeana is a specific case for the EU to boost everything.

MP: The national institutions will develop their own approach that works for them. There will be shared elements within institutions across the EU, but the approaches will always be inevitably tailored to the needs of the institutions. There may be a role for shared services internationally. [DRUID]. We have shared community tools that are used in parts of DP; and there will be more research and effort in that direction; the tools will become more mature; and institutions will get a better understanding of what they want out of such a service, e.g. a shared emulation hosting environment.

MS: There is a tension between EU funding and national funding. It works like a circle. At Sound&Vision there was a budget of 150 M euros for digitisation. Now we need money to keep the digital collection alive. That was not counted in. But the national government has no money; the EU can perhaps help in funding when the stakeholder area is larger.

- ***Is there reason to assume that the monitoring of long term preservation of digital heritage differs from that of other digital information (e.g. scientific data sets) and requires a separate methodology?***

MdN explains: most of our DP activities are related to institutional practices in collecting a specific type of traditional materials. With (born) digital, networked information things may change. E.g. the Library of Congress archive of Twitter messages is not a collection that can be compared with the physical collections with a national scope.

BS: The KB has collections that are not national per se, and they might relate to data that are elsewhere than in the Netherlands.

MdN: which government would be interested in the metrics about such a collection?

NG: Stresses that there is competition between institutions/nations. Comparing with all universities. JISC is set up for the UK higher education sector. More collaboration is easier with shared services, but competitiveness is at tension with that. UK universities are expected to be more globally competitive. The EU operates in a similar way on a more macro level. In universities there will be willingness to cooperate with other institutions on an operational level, but the management may want to be more competitive with another institution down the road. So an important question is: where is the benefit of collaborating with another institution?

MdN: Cultural heritage sector has different objectives and motivations, e.g. not per se to have an impact now, but also to have an impact fifty years from now: preserving the collections.

NG: The cultural heritage sector has different motivations.

MP: The definition of digital cultural heritage is changing all the time. Data sets may be cultural heritage in the future; scans of heritage monuments; content from the Web is/may be digital heritage as well.

MdN: In my view digital information, because we can keep it, is becoming more and more digital heritage. Think of Twitter tweets.

HV: There is a difference between monitoring in digital heritage institutions and other institutions, yes. Banks for instance have a different time perspective for keeping information.

MdN: But how does that result in a different methodology of measuring?

HV: Banks etc. use different standards. Comparing archives and museums with universities: universities have the latest information and technical knowledge, museums, especially small museums and archives they lack this technical knowledge, so these institutions need closer monitoring than universities.

MS: logically I don't think there is a difference between cultural heritage institutions and others. Keep it a long time for different reasons. It is about the value of the content (history, economy, commerce). The conception of cultural heritage changes. AV institutions now keep everything, but if that is going to be too expensive, they will drop some of the materials. Over time it will be about the value of the content, irrespective of whether it is cultural heritage or not. In the end also national governments will look at data and different values of data rather than at whether it is specifically heritage. At the Netherlands Institute for Sound and Vision

we keep all sorts of data, not only broadcast materials, because we have the infrastructure. And that data is kept at the same level of quality control as the heritage data.

MdN: But how does the institute account for that?

MS: There may be conflicts in business practices and government practices, I think. In 20 years' time there may be not a cultural heritage department and a scientific department, as we know it now.

NG: The idea of reuse might be similar across sectors.

- ***What are current monitoring DP activities that ENUMERATE should take into account?***

MS: A survey was done at the PrestoCentre (2012). Response was disappointing.

BS: The PARSE.Insight project (2010; it was a survey among scientists, librarians, publishers). Goal was to measure the status of DP in scientific research. It is publicly available.

MP: The APARSEN Survey - "What Is The State Of Digital Preservation In Your Organisation? - is doing similar things (costs, technology, digital library systems, business models). Work in progress, not available.

MP: ERPANET survey (2003-2004). Also case studies done with institutions and commercial institutions. It will be on the internet.

BS: ISO standard for metrics for web archiving (draft); questions related to preservation and webarchiving.

NG: JISC has not done things like this recently. There will be an evaluation of the impact of funding activities over the last 5 years, so a survey needs to be done.

NG: Another one is 4C [Collaboration to Clarify the Costs of Curation: http://4cproject.net/], focusing on costs and on costs determinants, strategies and tactics (EU project, FP 7, 13 partners). It will start 1st February. It is a 2-year project. We could compare our roadmap with ENUMERATE's. Focus is on what cost information can be used to make decisions. It is going to explore what's been done in the area. It's a coordination exercise. We don't want to come up with any new cost model.

MdN: Any monitoring or measuring in the archive world? What about APENET?

Survey in 2008, not much progress is being made in the past years.

MvE: Maybe the archival world is a bit survey tired. There is not much monitoring. Not on the APEX project as far as she knows. MvE will check what is done in the APEX project.

BS: Mentions DigCurv [http://www.digcur-education.org]. This project is about education and training of digital curation professionals and there was a survey on the topic. (DEN is involved in that project.)

NG: Neal Beagrie did a report in 2008 on DP policy issues. NG spoke to him recently. Beagrie thinks no big progress has been made since 2008.

MdN: Science, Libraries and AV archives seem ahead; Archives and Museums are perhaps a bit lagging behind. Questions are very general: Do you have a policy? Do you have an E-depot? Exception is the APARSEN project, this project tries not to focus on libraries only.

MdN: DP in museums still needs much attention, but the major asset of their collection databases - which have a very concrete value in museums - forces institutions to reflect on digital preservation.

MS: The Presto4U project [https://www.prestocentre.org/4u] by PrestoCentre builds 9 communities of practice within the AV domain: filmarchives, broadcast archives, teaching and learning repositories with AV materials, research organisations. We call it monitoring DP activities. It will be more qualitative than quantitative in the beginning. But the project is also building a methodology for making comparisons between mentioned communities of practice. ENUMERATE could be involved in the development of the methodology. Exchange can be made between this project and ENUMERATE – DEN. MS could be the contact person.

AR: NESTOR has done some survey work: http://www.langzeitarchivierung.de/Subsites/nestor/EN/Home/home_node.html.

AR: We had a questionnaire in PLANETS [http://www.planets-project.eu/docs/reports/planets-survey-analysis-report-dt11-d1.pdf] in which it is interesting to see which questions were misunderstood and in what way the answers were influenced by that. Sometimes I think that doing a meta-study could be more valuable than starting a new survey. Another suggestion would be to scan the archives of mailing/discussion lists for the term ´digital preservation´.

MdN asks to share information on other surveys initiatives should anyone touch upon potentially useful initiatives.

- ***Which of the specific standards for digital preservation (OAIS, PREMIS, DRAMBORA, TRAC etc) should ENUMERATE take into account? How can that best be done?***

MdN mentions some of the tools that are used in DP, wondering whether questions about such tools - often quite complicated and not easy to implement - should be part of a general DP survey. Would such questions as "Are you aware of this standard or tool?" be useful in a survey?

MS: There are industrial standards too. It would be useful to know whether (small) institutions are using flexible standards/tools or rather tools where they are locked in into some specific commercial product.

MdN: Some standards/tools are developed in the DP community and these are different from commercial/industrial standards developed for use in the market.

MS: But smaller organisations will presumably use commercial products, which imply the adoption of commercial (not flexible/open) standards. We'd like to know about this. NG: Be careful how you choose and present the list of resources/methods/tools, since the smaller organisations may not be able to answer specialist questions.

RG: It may be more worthwhile to ask for the principles behind the tools. Are these understood?

NG: Are you aware of checksums? etc.

MP: A problem is that if you are asking questions on this topic, you are implying that a specific standard has value for an organisation, but that is not always going to be the case.

BS: Small organisations starting with DP often start with looking at things like OAIS, because the big brothers are also working with it. Of course you must be careful not to suggest that if you don't use OAIS, you are not good in DP. But referring to these standards …I would do that. It gives recognition; they look at them so it must help them. PREMIS is a different case. A lot know PREMIS but few have it implemented; furthermore it is especially related to specific areas. In the libraries, it is used, but not in the archives. I would leave that out. But OAIS, DRAMBORA, RAC (successor to TRAC), these are generic standards, irrespective of specific kinds of collections. And even if the respondents don't know such standards, at least their attention is attracted. Second step is to focus on the principles behind the standards. The survey assumes that you have an image of DP. The survey should find out what that image is.

HV: I agree, OAIS is an overall framework. PREMIS could also be useful, mention it too. In a survey you could ask: do you have a solution for DP. The second question could be: "What kinds of standards do you use?" Also ask: "Do you have preference for open standards?"

Part II: Options for constructing a framework for collecting statistically-valid data

12.45 – 15.00 A review of questions on digital preservation policies and practice

MdN: Summarises the morning session discussions on the usefulness of finding a methodology for monitoring and measuring progress in DP across Europe, not just including the front runners, but including the whole field of digital cultural heritage. There are the three levels where metrics on DP were considered to be useful:

1. Internal use; internal accountability; measuring the progress that you made as compared with your strategies
2. Comparison among your kind (cost, investment, workflows, etc.)
3. Public accountability; what kind of metrics can help to make the case of DP with your government, the EU, etc.

AR: Will questions about outsourcing, subcontracting, evaluation of service providers and partnership be an issue in the plans of ENUMERATE? MdN: Yes, that is what ENUMERATE is interested in and what we will discuss this afternoon.

Options for constructing a framework for collecting statistically/valid data

In the second part of the meeting two topics are addressed:

   I.   *What kind of monitoring data are feasible***?**
   II. *To what extent can these data be quantitative in nature***?**

MdN: In the ENUMERATE Core Survey 1 (February 2012) on the topic of DP a question was posed: "Is your organisation included in an National Digital Preservation Strategy?" This caused some confusion among respondents, especially because it was not always clear what should be considered as a NDPS. In the follow-up ENUMERATE survey we will ask the national coordinators (representatives of the national governments in the activities of ENUMERATE) whether per country there is such a national initiative: "Is there a NDPS, endorsed by your government?". And if the answer is Yes: "Please provide a link to written material on the topic."

RG and HV then introduce the DEN/PACKED scoremodel, an online self-assessment tool for digital preservation. It is work in progress: http://www.scoremodel.org

The tool is based on OAIS and the Digital Preservation Capability Maturity Model (DPCMM) by Charles Dollar. The Charles Dollar model identifies 15 areas presented in a spider web. In the scoremodel developed by DEN/PACKED these 15 areas are organised into 7 high level categories:

1. Organisation (institution) and policy (does the preservation of digital files fit the structure and policy of your organisation?)

2. Preservation strategy (is it correctly recorded what is being preserved, for whom and in what manner?)

3. Expertise and organisation (of work) (is the right expertise present in your institution and is it put to good use?)

4. Storage management (is the physical storage of the digital files reliable?)

5. Ingest (are the right measures taken whenever a digital object is ingested into your storage system?)

6. Planning and control (is the management well prepared? Are all actions retraceable?)

7. Access (is access to the digital files properly regulated?)

AR: Are categories based on evidence? Do the answers have to be supported by evidence? Is it clear what is meant when an answer is given?

HV: It is intended for use by the organisation. It is not an auditing tool. It is low-level as far as the expertise needed to answer the questions is concerned. Detailed jargon about OAIS issues is avoided. It poses only questions in which Yes-No answers should be given, there is no weighing in the questions.

AR: There seems to be room for ambiguity in the issues addressed. A question about whether storage is stable, for instance, could lead to answers like: "Yes, it is stable; we have made a copy on an array system" and "Yes, it is stable; we have 5 array systems across 2 continents." You must try to be as concrete as possible. In a question like "Do you have sufficient policies?, you need to know when policies are sufficient. So you have to ask additional questions: " Can you provide an example?", or "Do you have some evidence?"

In the Review of questions on digital preservation policies and practices the focus group was asked to comment upon a set of questions based on/deduced from the Digital Preservation Maturity Model by DEN and PACKED. The DPMM questions merely served as a thread for the discussion. Is there something missing here? Are there any questions that you think are impractical or even nonsensical? Can we get proper answers? Is the terminology right? Below the topics and questions are presented in the column to the left; corrective comments and/or alternative wordings are in the right-hand column. Some general discussion points are summarized below each of the topics.

Questions

1. Organisation (institution) and policy	
1a. Is your organisation included in a National Digital Preservation Strategy? [Yes / No / Do not know]	- The question should be: "Is your organisation included in the national framework for digital preservation?" - Small institutions could be asked: "Is there an organisation/institution with knowledge on digital preservation which you can join/buy yourself into?" - Add here: "Please specify: …" - You might add: "Can you give a reference to the national strategy?", so at least you have some sort of evidence. - An overview of national strategies would be enormously valuable. - This is a question about organisation. Maybe more clearly make a difference between internal organisation and environmental organisation. (This is partly addressed in 2a.) - Be aware: technical and organisational strategy are very different things. Strategy may easily be related to migration and/or emulation by DP professionals, but that is not what is meant here. We are interested in something far broader than that at this point. 'National' is not really ok for Belgium; in Germany it should be 'federal'. - Maybe ask whether institutions feel there is a lack of national coordination, in whatever way? Or: whether institutions

	feel the need of such coordination.
1b. How is your contribution validated? [By law / By public agreement / Voluntary]	.

MP: Maybe take the discussion a step back: I'm not certain how you would interpret the answers to the first question (1a), how you would measure them. Which is best? Is it necessary to have a strategy, is it bad if there is no strategy? For some groups it may be a good thing to have a national strategy, for others it may not; it may not be the culture to work that way.

MdN: Is it necessary that there *is* a strategy?

BS: It may be too early to ask these questions.

MdN: Should there be involvement from the government in coordinating things like: responsibilities, preservation activities, shared storage facilities, etc. …for public institutions?

NG: It would be a good thing to have a government document somewhere, stating the responsibilities of heritage institutions as to the long term survival of digital heritage materials.

AR: Yes it is necessary to have a national strategy, since presently the situation is such that in several institutions staff members are still to be convinced of the necessity of DP. But it is very difficult to make progress here, to get it into cost line items or training programs, it is very hard to get that into an official, major component of their mission. If there was a national awareness about the need to establish a national strategy, that would be good for the institutions. Now the field consists almost entirely of projects and some ad hoc research.

MS: In Sweden, a couple of years ago, there was a bottom-up initiative of a couple of institutions - including the national library, but also smaller institutions - to organise DP. They called that a 'strategy', but it was not initiated by the government.

AR: Questions asked here are at a high level. What do we do with the outcomes of the questions? What do we do if someone is aware of auditing. What does 'awareness' mean? Examples: "Do you consider to organise auditing?" "How much would you invest in that?" "How many staff do you think you will need in order to prepare for such an audit?" In other words: which steps forwards can we make on the basis of these questions?

AR: Another problematic issue is that those questions cannot easily be answered by one person. If you can't fill out the questionnaire yourself - refer to the option: "I don't know" - maybe ask to give the email address of another person who can answer that particular question.

MdN: What ENUMERATE is trying to achieve is a baseline of core information about DP in Europe. We run a core survey with high-level questions that any institution with a curatorial mission should be able to answer. For the more insightful issues we organise a so-called Thematic Survey. Of course there are surveys exclusively focused on a particular community, like the PrestoCentre survey, but ENUMERATE is not set up to do that kind of research. A non-response, or a negative answer could still be informative if you want to have

a broad picture of what's going on in Europe. If answering the questions cannot be done by one person, and the questionnaire must be passed over to other persons in the organisation, the response rate will probably drop dramatically.

AR: But I think it is unrealistic to expect that one person can give real/satisfactory answers at all of these points that are addressed here. Unless he or she is very well informed about what's going on in the DP team.

MP: The skills involved are indeed various: technical skills, but also legal knowledge, etc.

AR: There is a tension between seeking detailed and generic information. We don't want 30 or so high-level questions that do not help us to do anything.

MdN: ENUMERATE is more about understanding the progress and less about best practices. In the cultural heritage landscape the majority of institutions is fairly small. We are not only interested in this 1% of frontrunners. We want to understand the big picture.

NG is concerned that specific circumstances may not be detected with general questions: if an organisation does not have a staff member who is responsible for DP, it could be that the organisation simply organises DP activities in a different manner. So actually you need to know more than what you get from some of these generic questions.

MS: Where do you want to go with your data? Who would you like to influence with the survey data? What is the purpose of the survey? We can make it very objective or make it more actionable.

MdN: Indeed we may consider asking questions that instead of being very objectively measurable, may lead to a picture of the intentions of institutions, e.g. by asking for opinions about what should be done in the area of DP.

GJN: There is a risk of getting socially desirable answers.

AR: Examples: "Yes, we need more training…" and then ask: "How much do you think your institutions would be willing to invest in training?" "Yes we need an aggregator." "Do you think your institution would contract a person for half a year to build that aggregator?" "No, because we don't have staff free for that."

NG: Everybody wants to be more mature than they are. In the AIDA project they used the Cornell maturity stages. It gives institutions a chance to map themselves onto some area. That way answers might be better confined.

MdN points to the three levels - discussed earlier - at which the survey could be useful. ENUMERATE is not only a top-down initiative (as NUMERIC was), it is also meant to help institutions improve their own practices. There is some pretentiousness here: there are certain activities/states of affairs in the area of DP, that may be needed or necessary, at least to achieve progress in DP (per institution). It would be great if there were three or four questions in the Core Survey, for which the answers given could be incorporated in the annual report of an institution (any), as a measure of the local state of affairs.

NG: I understand what you are saying, and I believe some sort of benchmarking approach might be valuable. We are trying to get some sort of snapshot of maturity. But the aim should

be to make that snapshot actionable, so that the next snapshot comes along. You might get that if you offer the option to nuance the answers: "No, but…" etc.

AR proposes to ask instead of very high level questions, not 200 low-level/detailed questions, but to search for some precise indicators, fine granular questions, that actually may serve as indicators of the high-level state of affairs. E.g. "Do you have someone who is responsible for managing the storage?" "Do you have someone who is responsible for taking a look at the file formats during ingesting?" "Do you have someone who is responsible for planning/training curating capacities?" Answers to these three questions actually can give a very clear picture of state of DP in an institution. It would be more informative, although perhaps the statistical validity needs close consideration.

NS thinks it could be a statistically valid approach to ask these low-level questions and then deduce high-level conclusions. It all depends on how representative is the way of combining the variables.

2. Preservation strategy	
2a. Does your organisation have a written Digital Preservation Strategy?	- The adjective 'written' is necessary here to get more meaningful answers. "Do you have a pointer to or at least a table of contents of that document, if you are not allowed to share the actual content?" …and "How often it is updated?" "How do you monitor your progress (self-assessment)?" This question can be put here, or in the section on Planning and control (6). - Extra options? "No, but we would like to have one…" or: "No, we are not interested in that…" - The question is essential. If you haven't got that, you have nothing you can build on. If you don't have that, then probably you don't have a strategy for migration (see 2b). - If respondents check the box Yes, you can ask them more advanced questions; if they don't you do not need to ask any further. - Beware of misunderstandings: a record management document might be misunderstood as a written DP strategy document.
2b. Do you have a specific budget for implementing digital preservation activities (e.g. conversion, migration, emulation, transcoding) [Yes / No / Do not know]	"Do you have a budget for implementing DP?" And then present the 4 or 5 topics (conversion etc.) separately. Ask for the order of magnitude per answer. - If you ask: "Do you have obsolete file formats?" then you can deduce from the

	answer given whether the institution has some sort of a monitoring system.
	• The questionnaire should be concrete. PDF/A is not obsolete, but is PDF?
	• Ask which formats are accepted in an institution´s repository.
	• "Is coordinating DP specifically a staff members' role?" …and then the answer could be "No, but we organise DP in a different way." And then you want to know: What way?
	• In the BL there definitely are activities in DP going on, but there is no specific budget (the budget is integrated, it is partly R&D budget for instance). The actual implementation is a cross team activity.
	• "Do you have a DP budget?" could be followed by "What is the budget spent on… ?" A follow-up question in the next survey could give you an idea of progress/change.
	• You do not need to ask about a budget if an institution answered no in question 2a.
2c. Is it part of your mission to select and collect born digital heritage? [Yes / No / Do not know] [If so: what types of born digital heritage?]	• You should be clear about your goal with this question. You should phrase the question with this goal clear in mind.

HV: Why is the question on born digital materials (2c) included here?

MdN: The question was included as a fairly low-level question, to find out whether the institution is faced with the challenge of DP. The urgency to do something in the area of DP with these types of materials is of course high. For digitised materials it is getting more and more common to throw away digital copies and digitise analogue materials later if it is requested, rather than turning to digital preservation. It may be cheaper. Digitising on demand. At this point the domain of DP has become a grey area.

MvE: Lots of money have been spent on digitisation. Often you cannot justify to just throw the digital collections away.

HV: Born digital is not so easy to define. A lot of digital culture is coming to us in the coming 10 years. It is a changing world. If you take a picture of a work of art you already have born digital materials.

MdN: Definitions is a daunting problem. A lot of what we are discussing is about definitions. To a certain extent you can manage the risk of misinterpretations. If 10% misunderstand the question, that may be balanced by the other 90%. In my opinion it is important to include questions about the way in which institutions deal with born digital materials, also because
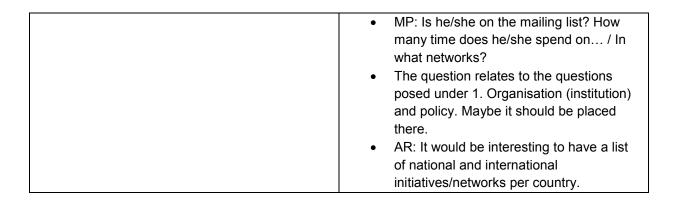
this may be a signal towards the institutions: there may be a challenge here. 50% of institutions of the Core Survey collect born digital.

MS proposes to design the survey in levels, where detailed questions are only asked if specific high-level answers are given.

MdN: Some questions might be rephrased: instead of "Do you have…" the question could be "Do you have a *need* for…". This would make the survey more actionable. So perhaps the focus could be more on the psychology than on the actual practice.

In general the conviction exists that you get a better understanding of the status of DP in an institution with lower level questions. We need very concrete indicators.

3. Expertise and organisation of work	
3a. Do you have a staff member whose tasks are specifically related to digital preservation? [Yes / No]	• When is someone DP staff? Is someone managing file storage DP staff? Surely file storage is part of DP, but the accounting system needs storage too… • Maybe focus on whether DP is managed? • If Yes: what do they do? What are his/her tasks? • If No: "Why not?" Ask for an explanation. • Or: "No, but we would like that." Or " No, but …" • "Do you have somebody who is responsible for training in digital curation capacities?" and "Do you have somebody who is responsible for digital preservation?" and "Do you have…". What are needed: low level questions used to create an indicator on the organisation of digital preservation on a higher level. • Instead of a question about the staff member, pose a question about an institution´s workflow here: "Do you have one or more workflows related to digital preservation?" [How many?] • Instead of a question on workflows, pose a question like: "How are the daily activities influenced by… [this goal of DP] • What kind of organisation do you think you are in? ("kind of" in terms of the maturity stages)
3b. If no: are digital preservation activities outsourced?	• A follow-up question could be: "What kind of activities?"
3c. If yes: Is this staff member involved in digital preservation networks at the national or international level?	• Question should be more specific! What can we learn from the answers given here?

	• MP: Is he/she on the mailing list? How many time does he/she spend on… / In what networks? • The question relates to the questions posed under 1. Organisation (institution) and policy. Maybe it should be placed there. • AR: It would be interesting to have a list of national and international initiatives/networks per country.

MdN: ENUMERATE tries to help institutions to improve their own work. Three questions on digital activities that every country could include in their national questionnaires. Help improve the knowledge and the discussion about metrics. Understanding your own practice and accountability to the government. We have sort of a missionary goal as well. We intend to point at strategic issues through these questions.

NG: If the aim is getting people towards strategies… On the whole there is not much about collaboration in the questions posed here. The intent to collaborate, for instance, could be related to the interoperability of systems. Answers might be: "No, we don't do that; but Yes we would like to collaborate."

MdN asks MP whether the metrics that the BL is now experimenting with are in some way comparable to the questions we talk about today. MP says that with the metrics at the BL are very clear, very focused. Does she have any examples? One of the strategic priorities is to ensure that the BL repository can store document collections for a long time… Another one of the strategic priorities is about collaboration, on a national and/or international scale. "It is important that we get the most out of our contribution to those projects, by implementing whatever we can internally." So one of the things being measured is the BL's success in matching those two things together, and this comes to looking at the number of tools or workflows that the BL has developed as part of these collaborative projects and that are implemented at home. It is quite specific.

GJN: Does this come to stating: if we co-developed 10 tools, then we will meet our criteria if say 5 of them were implemented?

MP: That is complicated by the fact that in certain projects the results may not be suitable for implementation at home. We've got about 3 or 4 things to measure for each strategic priority and there are 3 or 4 actions associated with each strategic priority.

GJN: Perhaps such specific indicators still could be used in other places as well. The BL's strategic priorities are not that specific and possibly a question about the number of workflows or about the number of tools implemented could be asked at other institutions as well.

MdN: A question about workflows - "Do you have one or more workflows in relation to DP?" - could be an indicator of whether your daily activities are influenced by the digital preservation goal.

MP: I don't know how useful that would be in terms of monitoring DP practices.

BS: It should be about whether you have implemented DP in your workflows. What you want to know is not whether an institution has written something on paper, but whether an institution's daily activities are tuned to the written document. You should have a question about how the daily activities are influenced by this DP goal/policy plan. AR: If you really want to know something about actual interest, you should also ask a second question otherwise everybody will answer that he / she is interested (social desirable answers) Are you interested in X? If Yes, ask: How much are you prepared to invest in this?

Do you have a need to participate in a national framework for digital preservation?

4. Storage management	
4a. How is the storage of your digital collections organised? [In house / Outsourced / Combined / Not]	• Be aware: some will tick all options (except the last one). • This must be quantified: perhaps ask to estimate percentages of storage per option? • You can only answer this question after having seen the answers. • There will be different answers for different (sub-) collections! So you may need supplementary questions. (Which is the case in 4b as well.) • Actually the question is about long term storage. • The last option is nonsense (Not at all).
4b. What storage media do you use? [Local server (LAN) / Networked server (WAN) / Cloud services / Offline magnetic tape / Offline optical discs / Other...]	• Suggestion: call it 'digital archive' instead of 'storage'. • Here again: some will tick all options. • Old and more recent solutions may co-exist. • Again: different collections can be involved, with different solutions for storage. • Are you looking for figures here? • An additional question could be about the amount of storage being managed (so add quantities!) If you don't do that, in 5 years time there will be no trace of change. (Proportionally there will be a shift in storage media.) • There is a danger in asking for numbers. Respondents will not know the actual numbers and just guess. • "What _long term_ storage media do you use?" • "What is your _preferred_ long-term storage? • Some reflection about asking for total amount of storage in use for DP purposes.

4c. Are you familiar with certification and auditing instruments for trusted digital repositories? [Yes / No]	• For many institutions this question will come too early… • …but in a few years' time all institutions will tick all (4c and 4d).
4d. Has your digital storage been validated through auditing or certification? [Yes / No / Do not know]	• Again: for many institution this question may come too early. The question may be too strict. Maybe it should be postponed. • The word 'validated' is too imposing. • There is no party that does auditing yet. • 'Digital storage' seems to suggest that this is about an institutions' infrastructure. The terminology should be adapted to 'digital archive'. • Have you performed/planned to evaluate according to the DSA [Data Seal of Approval: http://datasealofapproval.org/] set of …standards? • "Have you performed some kind of evaluation of your formal procedures?" • Maybe discern two levels here: (internal) self-assessment and formal auditing/certification by an external party.

5. Ingest	
5a. Do you add preservation information to the metadata of your digital objects? [Yes / No]	• Wording could be too specialist: "Do you add extra information to keep the information understandable?" • AR: Maybe ask to provide an example of their preservation metadata. • "Are you involved in collaboration initiatives to regulate ingest?"
5b. What formats do you accept in your digital storage? [Any native format / Any format after normalisation / Open standards only / Specific formats only / Other, ...]	• AR: be cautious. It is amazing to see how many formats are considered 'open' or 'standard', while they actually are not.

HV: What kind of organisation are you and what kind of digital materials do you have? You must have an idea of the complexity of the digital collections in an organisation. If it is only about an image database, it much easier to secure DP. In other words: how do you understand the composition of a digital collection?

6. Planning and Control	
6a. Do you have digital objects in your storage	• MS: What is storage in this context: the storage of masters? The storage in a

with obsolete file formats? [Yes / No / Do not know]	(dark) archive? Surely not the storage for immediate access on web server. • It should be about the storage of masters. • AR: Did you ever migrate objects? • MP: You could ask for the pro-active planning: "Do you have migration plans in place?"
6b. Do you have a strategy in place to deal with obsolete formats (migration, emulation etc.)? [Yes / No / Do not know]	• Difficult to answer! It assumes that organisations will be able to interrogate their entire store and find out what's in there; and then do an obsolescence check, but that is not a very straightforward thing to do. • Did you ever migrate digital heritage materials and did you experience problems with that? • Do you perform Quality Assurance on your digital and digitised content? What tools are you using?

MP: Is Quality Assurance incorporated in ENUMERATE's DP questions? I would be very interested to know how many institutions perform QA on their digitised content, before they ingest it into their repository. Which tools are you using?

MdN: I agree. It is essential. QA is not included in the Core Survey now. We shall look how we can include that point.

AR: Maybe ask some questions from the QA cycle, with verification at the end.

7. Access	
7a. Are your digital objects designated by persistent identifiers? [Yes, all digital objects / Yes, some digital objects / No / Do not know] (other questions on access are covered by another part of the ENUMERATE questionnaire)	• Add a question like: "How many objects can be presented to the public?" • Maybe a question about whether the objects will be accessible in an authentic manner or in the original context (or some representation of it). • Be aware of the fact that long-term storage can be organised very differently, with very different options for making the materials publicly accessible. (In a digital repository materials may not be readily available to the public.)

General remarks

AR: Is it possible to let different people fill out a part of the questionnaire. MdN: that is not the idea behind ENUMERATE. It must be possible for 1 person in the institution to fill out the complete questionnaire. Most organisations are relatively small institutions in which the director can fill out the questionnaire.

BS: Needed is a small phrase on what you want to get out of the questionnaire. Is it part of your strategy to collect born digital material? Phrase the goal of a question to explain, for the creators of the questionnaire (not for the creators). Make these ideas behind the questions more clear.

BS: There must be a possibility to print the questionnaire.

If you ask the following two questions: Do you have a fixed budget? AND Do you have an improved information plan? You can deduct other information of these questions. Such as no fixed budget and no information plan, than a lot will be ad hoc digitisation.

NG: [Maybe on the whole you should be more consistently] nudging institutions on their strategies. If you wish to collaborate, are you trying to design your process such that you can collaborate? This says something about your level of maturity. Or not at that stage yet but are interested in collaborating.

MP: Metrics should be very clear. Goals for instance are Collaboration national and international and get most out of contribution. Measuring the success of putting these two together. Working on tools that are not only of use for your own institution but also for other institutions.

HV: What about rights management?

BS: Presently there are no questions about rights; e.g. whether you are entitled to migrate things; to duplicate things; and whether you have agreements with the persons delivering the material; rights having to do with the use of an international E-depot.

> **Agreement: on the basis of this discussion on the questions we will reformulate the questions and send them to all participants of this meeting and ask them for their comments.**

Part III: A roadmap for future activities

NG: I will continue to monitor ENUMERATE and have a conversation when we are up and running. We might feed ENUMERATE pieces into the activities at 4C.

MP: We discussed a really interesting process and MP will compare the questions discussed today to the questions she is going to ask in the BL. They are at different levels, but that´s ok. Interested in seeing the final form before it goes out. She can ask colleagues, f.i. those working on APARSEN, for input.

AR: Very interested in what will come from this questionnaire. Questions can be used for further research or influencing. The more actionable recommendations we can deduce, if we have 5 pieces of evidence - we need more training etc. - that would be a huge argument. For instance more training or more R&D Scales. Arguments why to invest in these kind of projects (EU).

MvE: Very interesting discussion, but at the NA the focus is now a bit more on other things such as risks and making manageable the creation and selection process. And preparing the

NA for ingesting new digital materials. The focus is more on all digital information. We don't have the time and money to develop things ourselves, so at the NA we are interested in how can we collaborate with others.

BS: Interesting, likes to be involved in the rest of the process. I can use it internally and in the projects of the KB. Will give feedback.

HV: Thinks it is interesting for our own maturity model, to optimize this.

RG: The discussion yielded good input for the score model: http://www.scoremodel.org/

NS: It was interesting to hear the input of experts and input to optimize the questionnaire.

GJN: Has a bit of a headache from all the things that should now be taken into consideration. Especially determined to find the right person in an institution to fill out the questionnaire. Or get more people in an organisation to fill out parts of the questionnaire. The more you look at the indicators, the easier it will become to reduce the number of questions. Of course there always is a tension between top-down and bottom-up initiatives in survey development. Finally the aim is to have a questionnaire that leads to good feelings at both sides.