Questions for Portico and LOCKSS

As of 07 May 2014

1. Tech Questions
   1.1. Connection
      What must a participant do technically to participate?
   1.2. Storage
      1.2.1. Where are the data stored physically? 5.3.1
      1.2.2. How does your service structure its archival storage / archival packages? What is the granularity? 1.5.4
      1.2.3. To what extent is your system scaleable? Are there performance considerations? 1.4.3
   1.3. Ingest
      1.3.1. To what degree is the ingest process for content automated? 1.2.1
      1.3.2. How long does the ingest usually take from the point where the publisher agrees to make it available to when it is fully integrated into the system? How does this happen? Please describe the workflow and risks within the workflow. 1.2.2
      1.3.3. What technical requirements are there for the ingest process? 1.2.3
      1.3.4. Is the ingest limited to certain types of data (e.g. particular formats), and do certain types (formats) function better than others? 1.2.4
      1.3.5. Is metadata also ingested? If so, what metadata in what formats, and are there format limitations? 1.5.2
      1.3.6. Is there an audit trail (computer or human) for the ingest process? What quality control checks are performed? By whom and how much work is it? 1.2.5 & 1.2.6
      1.3.7. Are persistent identifiers ingested for each archival object and to what do they link? 4.5.3
   1.4. Security and Integrity
1.4.1. How do you ensure the integrity and authenticity of stored data? Have you proof? 4.5.2

1.4.2. How do you ensure security for your system and how to do you authorize access? Have you proof? 1.4.4

1.4.3. Do the technical requirements for authorization and identification require involvement by publishers and/or libraries? 1.4.4

1.4.4. What (other) measures ensure quality control? 1.2.6

1.4.5. To what degree is quality control automated? 1.2.6

1.4.6. What involvement by members or contractual partners is necessary to ensure the security, integrity, and authenticity of data? 1.4.4

1.5. Testing

The grant proposal requires comparison testing. How can project staff get access to run tests on your system? 1.7

1.6. Special services

To what extent can your system be extended? Can specific requirements by German libraries be built into your system? How? What about cases where the URL leads only to current content, or where a check-out process is necessary for a download, or where access tokens (sent by mail) are needed, or where a moving wall blocks access to either newer or older content? 1.4.2

1.7. Open Source

How much of your system is open source? 1.4.1

1.8. Long Term Archiving

1.8.1. What long term archiving services does your system offer?

1.8.2. How does your service handle format migration issues? 6.3.1

1.8.3. Can your system interact effectively with other long term archiving systems? 6.3.2

1.9. Disaster recovery

Have you evidence that your own service can recover from a catastrophe? 6.3.1

2. Process questions

2.1. Data exchange

2.1.1. What export formats does your service support? 1.4.1

2.1.2. Can your system exchange data with other archiving systems? Have you proof? 1.4.1

2.1.3. Can your system exchange data with library systems? Have you proof? 1.4.1

2.2. Trigger Events

2.2.1. Does your service provide access to post-cancellation content? Is it limited to a portion of the archived content? What implementation steps are necessary for
post-cancellation access? How long would it take from when this trigger event occurs to when the content becomes available to users? 4.4.1

2.2.2. How does your service handle a situation where one publisher transfers (sells) content to another? Is this situation provided for in agreements with the publishers? What implementation steps are necessary for this? How long would it take from when this trigger event occurs to when the content becomes available to users? 4.4.2

2.2.3. How does your service handle a situation where a publisher ceases operation? Would this apply to all content from that publisher? What implementation steps are necessary to address this situation? How long would it take from when this trigger event occurs to when the content becomes available to users? 4.4.3

2.2.4. How does your service handle a situation where a publisher's access mechanism has a catastrophic failure? Would this apply to all content from that publisher? What implementation steps are necessary to address this situation? How long would it take from when this trigger event occurs to when the content becomes available to users? 4.4.4

2.2.5. What participant involvement would be typical for the following scenarios: post cancellation, transfer, ceased operation, or catastrophic failure? Is content with very few users ("long tail") handled differently? Could a national consortium influence the tasks and user roles? 5.5.1

2.2.6. What is the interaction with publishers involved when a trigger-event occurs?

2.2.7. What mechanism is used to differentiate between a publisher's licensed content with permanent usage rights and archival rights and content without permanent rights? 4.4.1

2.2.8. How does a user become aware of a trigger event and how does the user get access in case of any of the trigger events? 7.1.1

2.2.9. When a trigger event takes place, how much would the look and feel of the content vary from that of the original? Would there be any loss of content or formatting? 7.1.2

2.3. Workflows

Can you provide organization charts or relevant diagrams to show the workflows, both automated and technical? Is there documentation for this and is it available? 6.2.2

3. Content questions

3.1. Content

3.1.1. Is a portion of the licensed contents in German libraries already available via your service? How much and how can it be compared? 4.4.1

3.1.2. Can you recommend how the extent of content in your archiving system could best be compared with the content in German libraries? 4.4.1

3.1.3. What mechanisms are used to ensure the completeness of content in your service's archive? Do these mechanisms rely on human assurances or on machine-based testing? 4.1.2
3.1.4. What is the average proportion of a journal’s content in the archive? What gaps exist and for how long? 4.1.2

3.1.5. How many eBooks and eJournals are in the archive? What about newspapers? 4.2

3.1.6. Can databases be archived? How many have been? What kinds of databases have been archived and has their functionality been preserved? 4.2.3

3.1.7. How has the amount of content in the archive changed over the last ten years? Per-year details would be welcome. 4.3.1

3.1.8. How can participants influence the selection of content for the archive? Content with a small (primarily German) audience is particularly at issue. 4.3.2

3.2. Metadata

3.2.1. What metadata formats are supported and what formats are used? 1.4.1

3.2.2. Do you alter publisher’s metadata? If yes, why and in what way? 4.5.1

3.2.3. Does your service have specific standards (e.g. metadata formats) for rights management? 1.4.1

3.2.4. Are there format limits for metadata?

4. Cost and business model questions

4.1. If a large consortium becomes a member or contractal partner, what options would it have to influence technological development and content acquisition and would this influence require (future) further financial contributions? 6.4

4.2. What are the cost models for your service? Are there up-front fees and/or ongoing fees? Are the fees scaled by size of library and are there discounts for consortia? 2.3

4.3. What would the ongoing costs be for a national consortium and how would they be calculated? 2.3

4.4. What costs would a Germany-wide consortium incur when its starts using your system? What are the direct and the indirect costs – that is, costs resulting from tasks the consortium itself must undertake. 2.1

4.5. Which costs incur when specialized German content is integrated into the service? (see special services above).

4.6. What models are typically used to divide expenses among the service provider, the individual participants (libraries and/or publishers) and a consortium? Personnel costs need to be part of this model. 2.1

4.7. Is the fee structure typically fixed during the contract period? If not, how does it vary? 2.5.3

4.8. What direct and indirect costs result when the following trigger events occur: 2.3.1

1. a library requests post-cancellation-access
2. a library requests access to content lost due to title-transfer
3. a publisher ceases operation
4. a catastrophic Failure at the publisher's server
5. a publisher can no longer host appropriately
6. specialized content using unconventional access methods becomes unavailable

4.9. Have the costs for your service changed in recent years? What might cause costs to change and are indications of change likely in the foreseeable future?

5. **Legal and licensing questions**

5.1. What are the legal usage options: 1) with no trigger event; 2) at a trigger event; 3) in case of post cancellation access? 5.3.2

5.2. What is the legal status of the organization or organizations that offer your archiving and storage services? 6.2.2

5.3. Please describe the structure of licensing agreements involving publishers, your service and libraries. What mechanisms are used to identify licensed content that can be made available to a library or consortium? What role would a (German) consortium need to play in clarifying the rights issues? 3.1/3.2

5.4. Which data protection and copyright laws apply to data ingested into your service? If the data and the access services are primarily in the US, does US law apply? Under what circumstances would German data protection and copyright law apply? What if the data were in another country (such as the Netherlands)? 3.3

5.5. What rights information for licensed content requires documentation and how is it handled for multiple institutions?

5.6. Where is the rights information stored? How can it be updated? 5.6.2/5.6.3

5.7. Are external rights management systems necessary and / or desirable?

6. **Participation and organization questions**

6.1. What type of cooperation with a German archive could you imagine? Would you provide contemporary examples? 5.2

6.2. What are typical formal conditions for membership in your organization or in organizations using resources that you supply? Who are typically the contracting parties and what legal decisions must be made? 5.1.1

6.3. What options are there for a national consortium to be a participant? 5.1.1

6.4. What is the usual period for such membership or contractual relationships, and how flexible is this? 6.2.3

6.5. What opportunities are there to be on the board of directors or relevant advisory committees? 5.1.1

6.6. Is there a problem for non-participants to get access when they have a legal right to?

7. **Sustainability questions**

7.1. How is your service financed? Are the operating costs completely covered? Have you proof? 6.2.1

7.2. How many staff are required to operate the service, what technical backgrounds do they require? What involvement is necessary by libraries or publishers? What happens if staffing (temporarily) falls below required levels? 6.2.2
7.3. What provision is there for maintaining service if you as a provider no longer can?
   What access rights to content persist if the service can no longer be provided? 6.2.3

7.4. Can a library renew membership / contractual relationships after a break and under what conditions? 6.2.3

7.5. What access rights persist if a library or a consortium ends its membership / contractual relationship? 6.2.3

7.6. How many libraries and publishers use the archival services and how has this developed over the last 10 years? How many German and European libraries and publishers take part? 6.1

7.7. Have there been external Audits in the past? When and by whom? Are the results available? Did the auditing authority issue any certification? Have any audit issues been addressed? 6.5.1

7.8. Did the audit include disaster recovery testing? 6.5.1

7.9. Are future Audits planned? 6.5.2