Central Commission for the Navigation of the Rhine

Consultancy on development and implementation of a harmonized database for test items for Boatmaster

Arnhem, December 2021

cito.com

Report CCNR

Consultancy on development and implementation of a harmonized date base for test items for Boatmaster

24 December 2021

Central Commission for the Navigation of the Rhine

Cito Team

Melchior de Vries, Caroline Jongkamp, Nard van Kessel, Servaas Frissen, Wil Knappers, Tecla Lampe, Margot Mol

Content

1 Summary

At the end of 2020, several Member States within CESNI have expressed interest in the development of a European multiple-choice database for theoretical examination of applicants who want to become a boatmaster. Cito was asked to find out what would be the best option for such a data-base and what it would mean for member states to implement this.

This study reflects this exploration and provides CESNI with an advise of the direction to take and the steps to be made, both in the short term as in the long term.

Through a series of interviews with stakeholders, we have captured the views member state representatives have of the current situation and the desired situation of the examination practice in member states. Based on these interviews, we can conclude that there is a broad support for a centralized database of questions which supports examination of candidate boatmasters. The advantages of such a database or item bank are well-acknowledged by all Member States. Representatives see that harmonization will lead to a higher quality of items and a more solid process of quality assurance. It is clear to all, that the investments needed for achieving this need can be modest. It is mostly time that Member states need to invest; time that will easily be paid back. At the same time it is clear that harmonization should not go any further than this – as for now. All Member States want to keep autonomy over their exams. Therefore, cooperation should be limited to building up a centralized database with questions which meet certain quality criteria. All Member States can contribute to this and all can benefit from it. Organizing such a data base should be done by an existing entity, as building a new entity would take significant effort and would distract energy away from the purpose of it. We believe that CESNI is the right organization for building up and maintaining the data base. We also believe it is good to keep costs low and have as much work done by existing staff as possible. Also, the investments in software and infrastructure should be limited. With a small and efficient organization it will be possible to produce results quickly and thus, take steps for future cooperation, which may go further dan sharing exam questions

In chapter 1, the background, purpose, research questions and methodology are presented. Chapter 2 contains the main findings, gained form the interviews and desk study. The conclusions, recommendations and follow-up actions on building and maintaining such a database of questions are presented in chapter 3.

Below, we sum up the answers of the main questions of this research and list the main recommendations.

Questions and answers

What are the advantages of the development and implementation of a European multiple-choice *database*?

Based on the interviews with stakeholders, as well as on our expertise in the industry, the following advantages were identified.

- Efficiency: developing questions centrally can save time and money, because member states need to develop fewer questions themselves when they make use of questions developed by others. The total workload is less than the current workload.
- Quality: a centralized database of questions can lead to a higher quality standard of exam questions, because all questions will go through quality checks that the member states have agreed upon.
- Quantity: a centrally developed database is most likely to contain more questions than the national databases, because the joint capacity to develop questions is larger than that of the individual member.

- Actuality: a common database is easier to maintain and is less likely to contain only outdated questions, because state-of-the-art procedures and supporting software will facilitate scheduled updates.
- Analysis and evaluation of items and exams can be organized centrally in addition to or instead of decentrally, with benefits of shared know-how, and opportunities for harmonized difficulty and reliability of the test.

What are the disadvantages of the development and implementation of a European multiple-choice *database?*

- Costs: selection of the software solution, setting it up, making procedures, collecting items and maintaining them requires an initial investment in time and money. The initial investment does not need to be large. There are good tools available at low costs. It is worth considering using software some member states have experience with.
- Usage: the time between creation and usage of items may be longer than the current practices in member states, because all questions need to pass the quality checks that the member states have agreed upon.
- Usage: the questions in the database may need to be in a common language for all, because translation and verification procedures are complex and costly in the context of high stakes exams[1]. Verification procedures are costly and will bring on issues with equivalence of the questions. Member states may have additional inconvenience due to need to translate from the source language into their own language of instruction. In order to minimize the impact of this, we recommend to start with English or German.
- Validity: measuring competencies is difficult with MC questions only, because a task where the right answer must be *recognized* may not be the best way to measure abilities to *apply* knowledge in a given context.

What are the main obligations for participating parties?

Participating member states will need to consider minimal obligations when participating in the central database. The obligations of each member state may differ, because the countries differ substantially in their ability to contribute, as well as in their intended use of the database. It is not realistic to assume that a member state with 10 exam takers per year will contribute as much as a member state with 500 exam takers per year. Therefore, CESNI member states must reach agreement on the conditions for participation, and on the differentiation rules in order to avoid obstacles to participate for some countries. There is one obligation that needs to be equal for all: each participant needs to endorse the quality standards and standardized procedures that the member states (will) have agreed upon. The standards and procedures should comply with best practice in assessment[2].

Participants can contribute in cash or in kind, depending on their 'reasonable share' in the effort. The member states all have experienced experts (some more than others) who are potentially able to develop

[1] Compare, for example, the translation and verification procedures applied in the OECD PISA studies.

[2] For further information, the reader is referred to the body of literature on assessment theory and practice. A standard textbook for the industry is Downing, S. M., & Haladyna, T. M. (Eds.). (2006). Handbook of test development. Lawrence Erlbaum Associates Publishers.

questions, to review questions, to adjust questions, or to join in a committee that approves items[3]. There is no reason not to benefit from this expertise. Contributions in kind by the participants will encourage the support base for the common effort, as well as give an impulse to the quality of the question development procedure. Participating member states may also contribute by submitting previously developed questions from their local database for approval and for sharing with the other member states. Last but not least, participants may contribute by taking part in work groups or project teams aimed at establishing and maintaining the technical, organizational and quality environment.
Each member state is encouraged to nominate a member (assessment expert) for the steering committee. The steering committee is an advisory group that provides strategic oversight and guidance to the project team. Also, each member state is encouraged to nominate (assessment) experts to join a working group. A working group is a group of resource persons who joins expertise in order to work on a specific subtheme or task during the project phase. These working groups can explore and prepare activities such as the contribution of questions from the national databases to the European database, choice of software, quality procedures, guidelines, etcetera.

Member states can choose their own path of participation, e.g.:

- participation level 1 = be informed through project updates, provide input for decision making about quality standards and procedures, and about conditions for use of the items.
- participation level 2 = level 1 + provide expert(s) for question review, for question approval, for sharing item bank / database know-how, for developing procedures, prepare for future use of the questions in the database.
- participation level 3 = level 2 + share items, share manuals, share quality assurance procedures, full use of the questions in the database.

How much would implementation of a centralized database cost?
The potential costs of implementing a centralized database are not easy to estimate. More information is needed to come to an accurate estimation. Factors that influence the cost are the desired time path, the chosen quality standards and procedures, the number of users[4], and the features of IT solution that CESNI is going to decide upon. A state of the art mature item banking system has relatively high costs in terms of annual licenses, but this may be outbalanced by relatively low costs of investments in the start-up phase. The cost estimates below give an indication of the cost range, from a build-your-on-solution to a high-secure, mature test item bank system with full workflow support.[5]

We recommend to go for the Basic scenario and focus on producing, sharing and using high quality content.

It must be noted here that the estimated annual costs are indicative only, and based on experiences with organizations[6] with similar objectives and approximately 30 licensed users of the software.

---

[3] Source: interviews with stakeholders. See annex 3.

[4] The license costs for commercial software as well as support costs often depend on the number of persons who are allowed to use the software.

[5] The reader is referred to chapter 4 of this report for further details and guidelines for selecting a solution that fits the needs of CESNI.

[6] E.g. Cito, clients of Cito

Table: Cost scenarios for the first three years.

Scenario	Content quality standards	IT features	estimated annual cost
Basic	Relatively low standards for quality assurance, such as screening procedures and approval. Simple procedures.	Basic build-your-own-solution, e.g. based on SharePoint lists, document libraries[7] or Office 365 applications	• Manhours; approximately appr. 1500 - 3000 euro per year • IT solution: 1.000-2.000 euro per year
Mature	High quality standards with regard to authoring procedures, question life cycle maintenance, approval.	Third party itembanking system with full authoring workflow support, advanced security guarantees, rich reporting, and compliant with qti industry standards	• Manhours; approximately 5.000-10.000 euro per year • IT solution: 50.000-80.000 euro per year

What is the realistic timetable for implementing a centralized database of test questions?
We believe that a timeline of three years is realistic in order for CESNI to implement the common European database. The situation after three years can be described as a fully operational database[8] in place, accessible to all member states who choose to use it. For this to be realized successfully, a number of actions need to be taken and decisions need to be made, which are outlined in more detail in chapter 4. One of the conditions for success is to have a small but dedicated and skilled project team working on the labor-intensive first three years. In the first period, a plan must be made, which contains a description of the scope, the quality criteria, the procedures for quality assurance, the conditions for adding and retrieving items, the contact persons from each Member State, the project organization and the project planning, a description of the tooling that will be used etc.

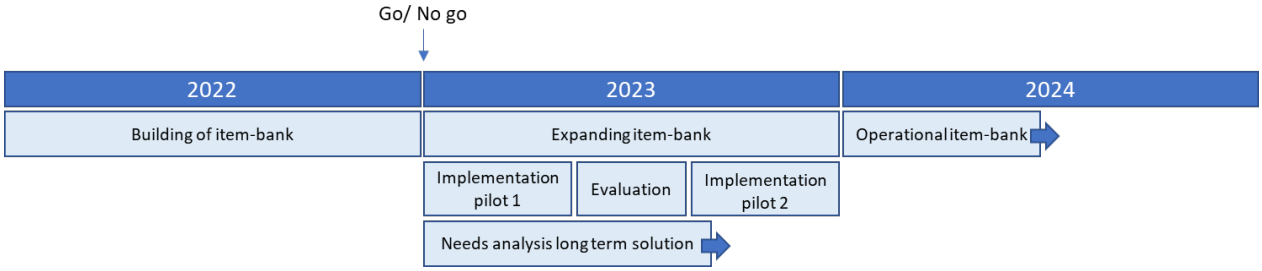
Member States will need to agree to the plan as their formal support is needed for the success of the project.

---

7 More details on the Sharepoint lists and Sharepoint libraries can be found in Annex 6 of this report, and on the website of Microsoft (Sharepoint List https://support.microsoft.com/en-us/office/introduction-to-lists-0a1c3ace-def0-44af-b225-cfa8d92c52d7; Sharepoint Library https://support.microsoft.com/en-us/office/what-is-a-document-library-3b5976dd-65cf-4c9e-bf5a-713c10ca2872; https://support.microsoft.com/en-us/office/introduction-to-libraries-7d4221d9-8fb9-40d5-8441-2374c84b5e26)

8 The quantity of questions at this point, will of course depend on many factors to be decided. For example, we have assumed (see chapter 4) that at least some of the member states will be willing to make questions available from their own database in order to get a good headstart. The conditions to make this possible will of course be subject to decision making and negotiation.

After the plan has been approved of, the activities that lead to a first pilot can start. The pilot is necessary for checking procedures and standards, but also for making member states enthusiastic.



The reader is referred to chapter 4 for more detailed recommendations on the set-up and implementation of the database.

In the table on the next page, the main recommendations with regard to how, when, what and why are listed. For more elaborate clarifications, the reader is referred to chapter 4, section 4.3.

Getting started

The upcoming phase of this project will be one of further decision making and aligning interests. The process of coming to this report has been an early and necessary step in that phase. It is now up to the stakeholders to keep the momentum going. The goal of having an elementary item bank in place within 2,5 – 3 years is realistic.

We therefore wish to encourage CESNI to:

- endorse the recommendations in this report
- take decisions on several points indicated in this report
- assign a project team to start building the item bank.

Recommendations

	Aspect	Research questions		Recommendations	Evidence
0	Scope / phasing		0.1	We recommend to keep the scope small in the first stage and build an easy-to-implement database, filled with exam questions from all participating member states. The main reasons are the preference as expressed by stakeholders for quick tangible results for the database as such, and importance of autonomy for member states when it concerns test assembly and test administration.	Interviews
			0.2	Make a detailed project planning and take approximately 2-3 years for full implementation. Based on best practice with projects of similar scale, this is the time commonly needed to build an operational item bank.	Assessment best practice
1	Organization	Question 1.1, 1.2, 1.3	1.1	Make CESNI responsible for the content, the operation and the maintenance of the database. This option has ample support from the member states; any other option is not logical.	Interviews
		Question 1.4	1.2	Formulate quality standards and procedural conditions for member states to contribute items to the new international database. Create review and an approval procedures. Formulate conditions for member states for extracting items from the item bank.	Assessment best practice Assessment best practice Analysis
		Questions 1.5, 1.6, 1.7, 1.8	1.3	Make a (small dedicated) project team responsible for developing procedures and guidelines.	Analysis
2	Technical	Question 2.1	2.1	Ensure that the database enables import, export and publication of items. These functions will do for the first phase.	Assessment best practice

	Aspect	Research questions		Recommendations	Evidence
		Question 2.2, 2.3, 2.4	2.2	The database should contain basic functions, such as select and sort, which are needed for efficient retrieval of questions.	Assessment best practice
3	Content	Question 3.1, 3.2	3.1	Assign a person who is responsible for the definition of metadata which characterizes item life cycle and re-use. This task requires specialized skills and expertise.	Assessment best practice, Interviews
		Question 3.3, 3.4	3.2	Accept only items that have passed a thorough quality review procedure, because joint quality assurance is one of the main factors that will make members enthusiastic.	Assessment best practice
		Question 3.5	3.3	Collect, maintain and provide the items in one language only, and, for mainly practical reasons, we suggest that language to be English or German	Analysis
		Question 3.6	3.4	Use multiple choice questions only in the short term, but keep options open for other item types in the long term. This appears to be acceptable to stakeholders and reduces costs and complexity in the short run.	Interviews
4	IT-aspects	Question 4.1, 4.2	4.1	Conduct a brief study for the selection of the software. Consult member states who have experience with this. Make a decision after comparing the pros and cons of available options.	Assessment best practice, Interviews
		Question 4.3	4.2	Decide whether data may or may not be stored in the cloud.	Assessment best practice
5	Interests of participants	Question 5.1, 5.2	5.1	Make sure that each member can benefit from the new database; it should not be a step back for any of the member states.	Interviews
6	Security	Question 6.1	6.1	Choose for a solid and reliable authentication procedure for all users, after comparing the pros and cons of each option.	Assessment best practice

	Aspect	Research questions		Recommendations	Evidence
		Question 6.2	6.2	Use procedures for data registration which are compliant with GDPR (EU-General Data Protection Regulation), in order to comply with laws on storing personal data.	Assessment best practice
7	Support	Question 7.1	7.1	Formally invite CBR and up to three other 'frontrunner' organizations to share (A) part of their item bank, and (B) their know-how, procedures and manuals on the authoring and item banking process. A group of four frontrunner organizations may be considered a sufficiently large 'critical mass'. The project will then have a solid foundation to build upon and to grow.	Interviews
		Question 7.2	7.2	Consider involving external support during start-up of the item bank, if CESNI would require some guidance in the next steps towards building a question database..	Assessment best practice

2 Research design

In the following paragraphs, we describe the backgrounds, reason, purpose, research design, the main research questions and the methodology of this report.

2.1 Background of the assignment

As of 18 January 2022, a new legal framework based on competences in European inland navigation will be applied. Hence, a harmonized examination of competence requirements throughout Europe is desirable. This will contribute to safety of inland navigation, as the required competences can be demonstrated by applicants all over Europe.

For this, the European Standard for Qualifications in Inland Navigation (ES-QIN) establishes a European framework of references for qualified personnel in inland navigation. The framework adopts, in a standardized way, the requirements of Directive (EU) 2017/2397. ES-QIN 2019/1, containing provisions on competences, practical examinations, approval of simulators, medical fitness and models of crew-related documents. Tied to the framework are harmonized competence requirements for practical examinations and for theoretical examinations for boatmasters.

At the end of 2020, several Member States within CESNI have expressed their interest in the development of a European multiple-choice database for theoretical examination of applicants who want to become a boatmaster. Member States of CESNI want to explore options towards a truly European database for the examination of applicants who wish to obtain a certificate of qualification based on the ES-QIN.

Implementing a European multiple-choice database for examination is a complex task, as it involves besides political will, technical aspects, IT-aspects, pan-European organizational aspects as well as matters related to content and test development.

2.2 Reason, purpose and research design

Reason of the assignment
CCNR is seeking for an advice on the development of a European multiple-choice database for examination of applicants who want to become a boatmaster.

Purpose of the assignment
The consultancy should lead to:

- Clearly identified advantages of the development and implementation of a European multiple-choice database;
- An overview of possible disadvantages of such a database;
- A clear description of the obligations of participating parties, both for countries with many applicants per year as countries with only a few applicants per year.

In order to serve CCNR with such an advise, we conduct an analysis that contains all necessary elements for the Member States to decide whether they want to participate or not. The advice also includes elements that Member States may trigger to participate. The analysis includes all questions listed in the Terms of Reference (ToR).

Research design
The research design is divided into three parts, which we describe below**.**

Desk research
Before we started with the interviews, we have studied relevant documents, in order to be sure not to waste any time on matters that are clear already, but also to ensure all relevant topics are covered in the research phase. The sources of documents were those provided by the steering group as well as those found through internet search, for example on websites and online archives of EU and of the commissions for Rhine, Danube and Sava river[9].

Interviews
CCNR proposed representative relevant stakeholders, both from CESNI member states that are frontrunners as well as from members states that are not, or who's interests and stakes are simply smaller. The interviews were conducted by experienced consultants who have in depth knowledge of exam systems and databases.

Analysis and reporting
We combined the insights from the desk research and the interviews with the expertise of Cito database experts. This has led to a report that must help decision making. It contains the following topics:

- Advantages and disadvantages of the development and implementation of a European multiple-choice database.
- Obligations of participating parties.
- Cost estimation, including costs of developing the database as well as implementation and maintenance.
- A quality management plan, including an overview of all relevant content matters for exam making and exam taking.
- An indicative timetable for development and deployment.

We presented the intermediate results to CESNI/QP on 16 September. On 1 October, we had a follow-up discussion with the steering group of CCNR. The additional information gained in this discussion was used for completing the draft report, which was shared with the steering group. The steering group met again on 5 November. The feedback on the draft report leads to a round of revision and delivery of the final report.

[9] Examples of relevant sources and documents (not exhaustive):

DIRECTIVE (EU) 2017/2397 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 12 December 2017on the recognition of professional qualifications in inland navigation and repealing Council Directives 91/672/EEC and 96/50/EC: OJ L 345, 27.12.2017, p. 53–86.

https://www.cesni.eu/wp-content/uploads/2020/03/ES-QIN_2019_en.pdf

http://www.danubecommission.org/uploads/doc/Presentations/2016/2016%2012%2009%20Danube%20Commission%20Plenary%20Presentation_Final_DO_DT.pdf

2.3 Research question and methodology

Research question

Based on the Terms of references (ToR) the *main question* for this assignment was: **how should a European multiple-choice database be developed, for examination of applicants who want to become a boatmaster**?

Relevant *sub-questions* are:

1. What are the advantages and disadvantages of the development and implementation of a European multiple-choice database?
2. What will be the obligations of participating parties?
3. What is a realistic cost estimation, including costs of developing the database as well as implementation and maintenance?
4. What should be the quality management plan, including an overview of all relevant content matters for exam making and exam taking?
5. What could be an indicative timetable for development and deployment?

Methodology

By structuring the questions from the ToR, we noticed that some of the areas needed for solid and substantiated decision making are covered, but not all. Below, we have listed the questions from the ToR by topic and we have added questions we find relevant for the analysis. In the first part of the project, we have studied this in more detail.

Topics and questions to include in the research

<u>1 Organizational aspects</u>

1 Which organization or body should and could become responsible for an international database?
2 Who should be responsible for the content? Who for operation and maintenance?
3 How should this be formally organized?
4 How can transition in terms of responsibility for content be organized between the contributing national database(s)?
5 How should this be practically organized?
6 Who has access to the database / item bank?
7 Item bank ownership; what roles and responsibilities must be defined?
8 Availability and access to documentation such as manuals, user guides, instructions, checklists for all members involved.

<u>2 Technical aspects</u>

To what extent should the database:

1 Consolidate all aspects of exam management into one system with a single view, enabling access, control, sort, analyze and export items into an exam delivery system?
2 Provide a centralized managed repository for items?
3 Provide an exam delivery system for candidates?
4 Provide custom data analytics and reports?

3 Content aspects

1 What should the database consist of? Is it merely a collection of questions and answers? Or also software to prepare an examination, choosing from the pool of questions the set of questions put to a candidate?
2 How should the database be set up and maintained regarding content? (Item life cycle; re-use of items).
3 What could be done to make sure that pedagogic / educational expertise is applied when elaborating and evaluating the questions and answers for the database?
4 How can the quality be ensured? In terms of quality of content, quality of translations, psychometric quality.
5 In how many languages (and which) should the exams be available?
6 Should there be any additional item types other than MC?

4 IT aspects

1 How should the database be set up and maintained, regarding IT?
2 What logical options are there for setting up a database that serves both short term goals and long term goals?
3 What kind of hardware is needed to appropriately operate the software?

5 Interests of participants

1 What could make countries enthusiastic about or what could discourage them participating in a European database?
2 Will the central database provide added value to Member States, for example if they can focus resources on content of questions instead of carrying out [supervising] examinations?

6 Security

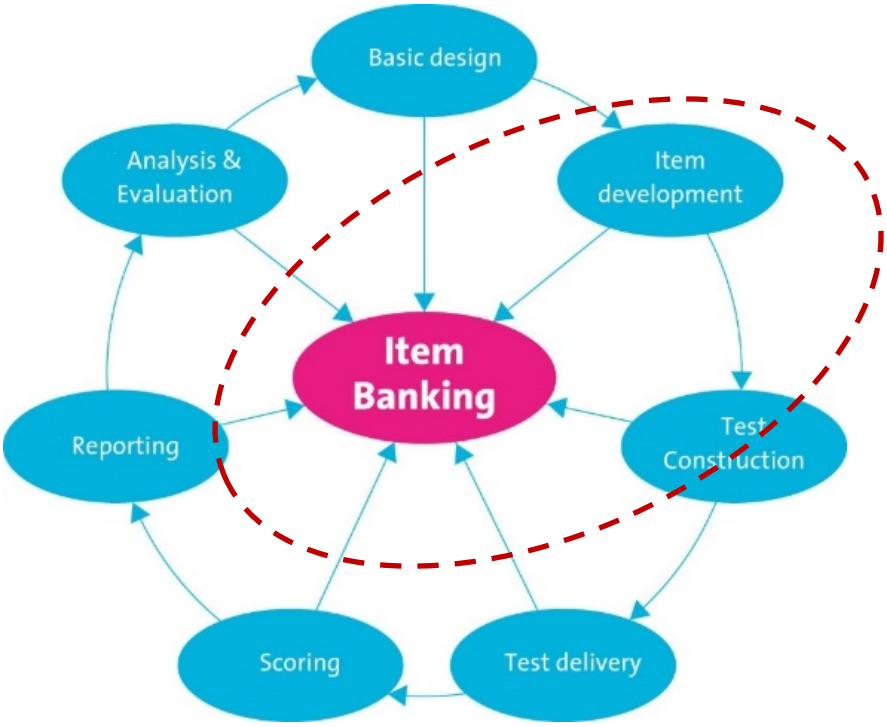
1 How can confidentiality be ensured?
2 What privacy matters need consideration?

7 Support

1 Which contribution can be expected in the start-up phase from organizations such as CBR?
2 What support is needed, or worth considering for sustainable database management after delivery?

A structured approach

We use the ‘test cycle’ as a guiding model for information gathering and analysis. The test cycle contains all elements of exam making and exam taking that are linked to item banking, which is the activity of building and maintaining a database of questions. Using this model ensures that no vital aspects are overseen in the investigation.

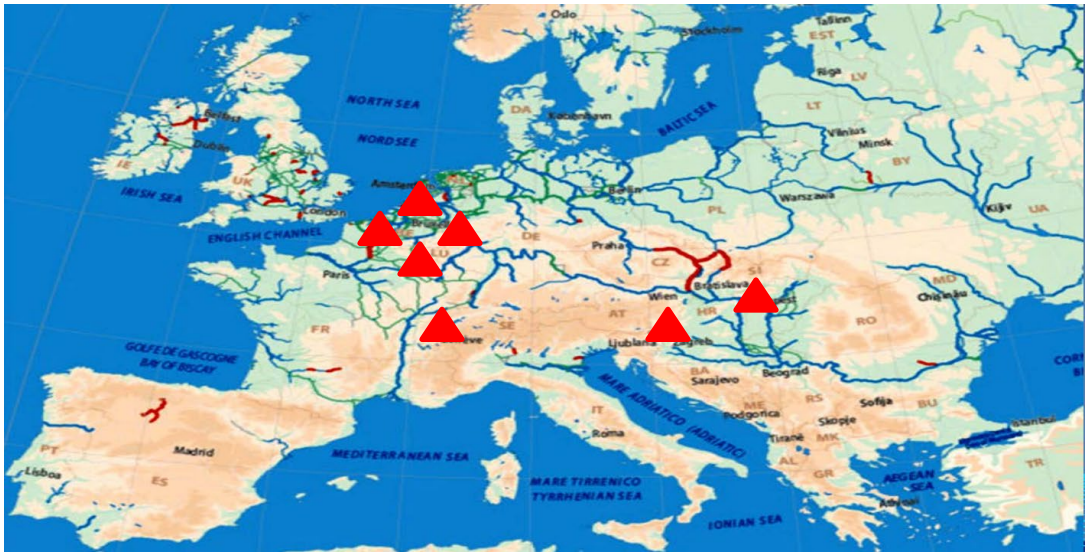


In the course of the research, the focus converged to the relation between item development and item banking. Of course, the relation with other elements of the test cycle were never out of sight but had less focus because the steering group and the stakeholders expressed that these elements should have a lower priority as compared to the more urgent questions of building, filling and maintaining the item bank.

3 Findings and analyses

We have held eight interviews with stakeholders from different member states. These interviews took place between August 18 and October 15, 2021. The stakeholders were based in member states Germany, Belgium, the Netherlands, Luxemburg, Switzerland, Croatia and Hungary (see figure). The commissions that these stakeholders together represented were the commissions for the Rhine, The Danube and the Sava.

Figure: the inland waterways in Europe and location of the stakeholders that were interviewed

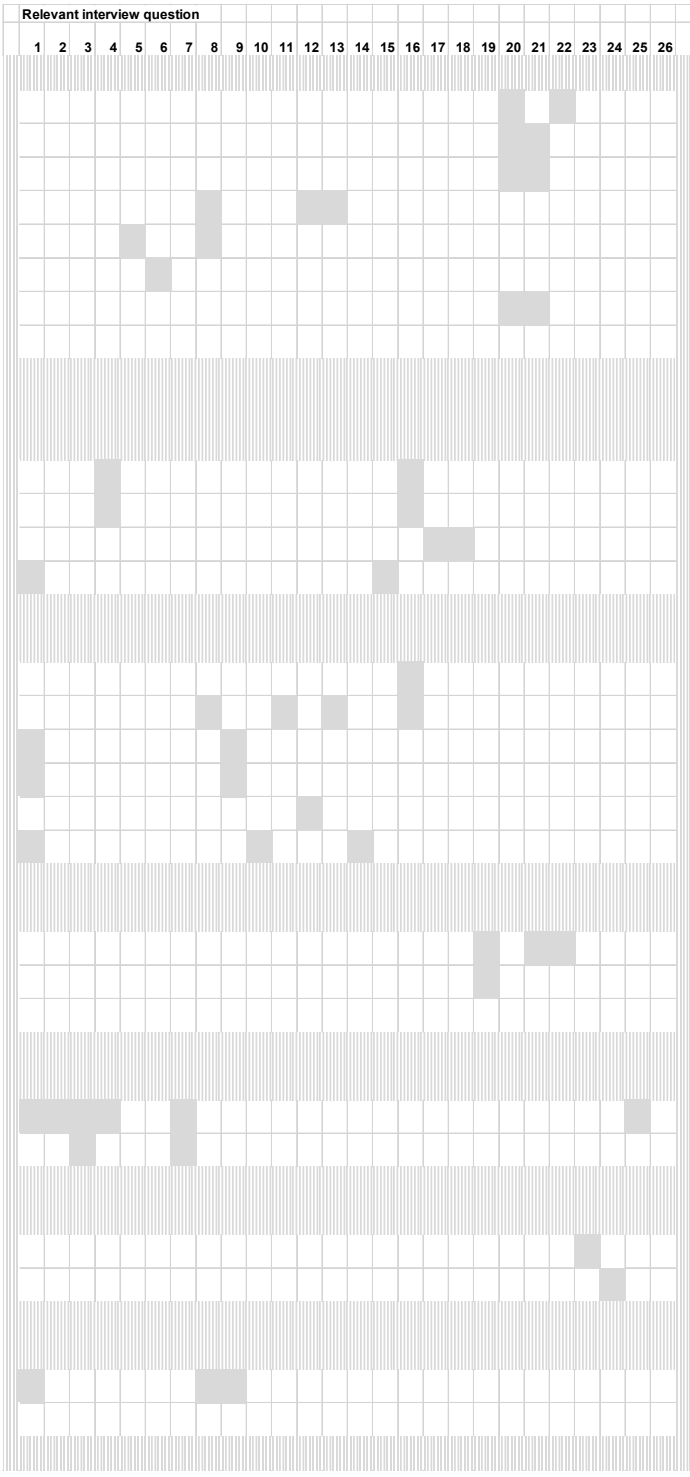


Source of the map: D. Oen, D. Theologitis, and B. Urrutia. “Inland navigation developments in the EU 2016” Presentation for the Eighty-seventh session of the Danube Commission, Budapest, Wednesday, the 14thof December 2016. European Commission -DG MOVE http://www.danubecommission.org/

The interviews were structured in a way that first the current situation of examination was described (the ‘ist’ situation), and next the views of the respondent on the future situations (the ‘soll’ situation) were explored. The views on the future situation were described according to the questions that were listed in section 2.3, and repeated in the table below. The list op 26 interview questions can be found in Annex 1. The interviews were semi-structured, in order that the stakeholders could speak freely nd share the topics and views that they wanted to bring forward. In Annex 2, the summary notes of the interview responses are found.

The table on the next page summarizes the links between the research questions and the interview questions. Of course, the interviews did not provide answers to all the questions, therefore information was added that originates from what Cito has experienced and learned from best practice and industry standards in the field of assessment and the application of test item databases.

Table: links between research questions and interview questions.



Topics and questions to include in the research
1 Organizational aspects
1 Which organization or body should and could become responsible for an international database?
2 Who should be responsible for the content? Who for operation and maintenance?
3 How should this be formally organized?
4 How can transition in terms of responsibility for content be organized between the contributing national database(s)?
5 How should this be practically organized?
6 Who has access to the database / item bank?
7 Itembank ownership; what roles and responsibilities must be defined?
8 Availability and access to documentation such as manuals, user guides, instructions, checklists for all members involved.
2 Technical aspects
To what extent should the database:
1 Consolidate all aspects of exam management into one system with a single view, enabling access, control, sort, analyse and export items into an exam delivery system?
2 Provide a centralized managed repository for items?
3 Provide an exam delivery system for candidates?
4 Provide custom data analytics and reports?
3 Content aspects
1 What should the database consist of? Is it merely a collection of questions and answers? Or also software to prepare an examination, choosing from the pool of questions the set of
2 How should the database be set up and maintained regarding content? (Item life cycle; re-use of items).
3 What could be done to make sure that pedagogic / educational expertise is applied when elaborating and evaluating the questions and answers for the database?
4 How can the quality be ensured? In terms of quality of content, quality of translations, psychometric quality.
5 In how many languages (and which) should the exams be available?
6 Should there be any additional item types other than MC?
4 IT aspects
1 How should the database be set up and maintained, regarding IT?
2 What logical options are there for setting up a database that serves both short term goals and long term goals?
3 What kind of hardware is needed to appropriately operate the software?
5 Interests of participants
1 What could make countries enthusiastic about or what could discourage them participating in a European database?
2 Will the central database provide added value to Member States, for example if they can focus resources on content of questions instead of carrying out [supervising] examinations?
6 Security
1 How can confidentiality be ensured?
2 What privacy matters need consideration?
7 Support
1 Which contribution can be expected in the start-up phase from organizations such as CBR?
2 What support is needed, or worth considering for sustainable database management after delivery?

3.1 Organizational aspects

3.1.1 Current situation of examination

We asked about the current situation of examination, in order to find out what the main differences are in the way examination takes in different member states and establish the starting point for each country.

- Exam creation differs from local experts who create exam questions to standardized procedures for creation and approval
- Some member states have outdated question catalogues whereas others have regular updates of questions, based on new European standards, test grids, psychometric analysis
- In some member states test taking takes place orally. In others written exams are the standard. We also saw large differences in the type of questions. These varied from essay questions and assignments to multiple choice only and fully digital.
- Transparency/confidentiality: public catalogue of questions *vs* 'black box'
- Reporting on results: live by examiner *vs* automated digital

3.1.2 Pros and cons of current practice

We wanted to know how stakeholders perceive the current practice of examination. The following pros of the current practice were mentioned. We consider this useful for discussions about the extent to which content and processes should be harmonized.

In countries where oral examination take place, advantages stated by representatives were flexibility, individual customization and the possibility to resolve misunderstandings when needed.
Different interviewees said that they consider it valuable that candidates get the results directly after the exam. This is possible both in oral examinations and in fully digital examinations.
Some stated that they value autonomy they have in item and test construction.
More specifically, the representative from CBR mentioned that the item bank with MC questions is fraud-proof, and allows good analysis of exams and assignments.

In the interviews it became clear that the current examination practice in some countries faces problems that need to be solved. Some members states have an old and outdated catalogue of questions as well as an old and outdated examination system. Also common practice: the quality of examination depends on quality of examiners; Comparison between exams virtually impossible.
In some countries the handling time per candidate is simply too long. Others stated the high costs of their exam system.
To sum it up, each country has its own needs for improving the examination practice.

3.1.3 Benefits of and obstacles for a centralized database

Benefits

Next, we asked about potential benefits of a centralized test database.

- Uniformity: harmonized way of examining candidates
- Efficiency: questions would be developed centrally; thus resulting in higher efficiency
- Quality: items will be created and selected according to higher standards.
- Standard way of exam taking will ease the process for examiners
- Actuality of questions will be ensured better

- Costs can be shared amongst participants
- Work load can be shared; participating countries can focus better on non-generic components

3.1.4 Obstacles

The stakeholders mentioned the following obstacles for the use of a centralized test database:

- It may require a lot of work and considerable investments
- One central organization is needed for the management
- Tailoring exams to regional needs may be more difficult
- Less flexibility; slower response on required adjustments; these will be directed centrally.
- Language can become an issue for some candidates

3.1.5 Leading principles when developing a centralized database

The stakeholders mentioned a number of leading principles:

- Organization: It must be clear who is responsible, one central administrator. A well organized secretariat is necessary. In case of problems in questions, fast reaction and solution is required.
- Content: Many experts are now working on it, specialists, that must be preserved. Validation is important.
- Content: It is important to guarantee that the database is constantly kept up to date and is therefore immediately adapted in the event of changes to the law or new rules. Harmonization is also important: everyone should use it.
- Stakeholders: National and regional differences must be preserved. The countries' own influence and contribution are important in the case of a central database and must be properly regulated. The countries need to have a say in the quality of the exercises and the exams.
- Technology: Preferably a web-based solution
- Security must be assured.

3.1.6 Contributing to the database

All stakeholders said they want to contribute to the database by means of knowledge and helping hands to develop items, review items and adjust items.
Interviewees showed little eagerness, however to investing money in it.

3.1.7 Responsibilities

We asked about the views of stakeholders on responsibilities for the database. According to the interviewees, responsible for content could be:

- CESNI
- A neutral body, i.e. a third party, not a stakeholder; a strong and rich organization or entity
- No organization, but merely cooperation of the examination boards

Responsible for the functioning of the database should be:

- EDINNA (mentioned once)
- Neutral body / one central organization
- All member states; no separate organization or institution

3.1.8 Test taking

In order to arrive at a uniform central database, or to any other form of more intensive cooperation, it is good to realize that the stakes differ among member states. This is expressed best in test taking. Here we see big differences, for example in the number of candidates. These vary from 40 per year to a maximum of 500 per year. Of in the number of locations. In small member states there is one locations, whereas there is also a member state with 7 examination locations.
Variety also exists in moments candidates can take exams. In some states there are fixed dates, in others dates are flexible and depend on the availability of the examiners.
The differences in the current practice of exam taking influences the opportunities for setting up and implementing an exam system. They do not affect the possible implementation of a database.

3.2 Technical aspects

3.2.1 Scope: database system versus exam delivery system

In some of the interviews, the long term perspective of a complete centrally managed exam delivery system was discussed. For now, there is not enough support for such tooling among the people we interviewed.

3.2.2 Data analytics and reports

Most stakeholders did not have a clear view on the way the database can support data analytics. Therefore, we think this is a topic for discussion in a later phase. For now, it would be good, to take data analysis into account in choosing appropriate database tooling.

3.3 Content aspects

3.3.1 Content domains

When working towards a uniform database, we think it is good to have a notion of the most common topics to cover. These are the one which were mentioned most:

1. EU legislation
2. ES-QIN competencies as established by CESNI
3. The police rules and European rules of the different Member States
4. All theoretical examinations on competences for the inland skipper

Interviewees expect up to 20% of the exam content to be country-specific.

3.3.2 Item types

Looking ahead to what stakeholders expectations are about the type of content of the database, these were the items types mentioned.

1. Multiple choice, with photos, illustrations and schemes,
2. Multiple choice with multi media assets such as video and animation
3. In future also drag&drop, hotspot, yes/no, fill-in-the-blank, sequence questions,
4. Open questions using features such as drawing and simulations
5. Item types which support measuring competencies and skills

We see a phasing in the type of questions is needed first project phase and what types can be added later.

3.3.3 Versions

As for the number of questions a database should contain, it is important to know how many versions of an exam stakeholders expect. Most representatives expect up to a maximum of 4 versions per exam. Others preferred an infinite number of versions as they foresee exams to be created out of a large item bank.

One stakeholder emphasized that, regardless of the number of test versions that could be made, it is for each member state important to continue to compile own exam versions.

3.3.4 Languages

Stakeholders have the following expectations about the languages that should be available in a test platform. This is relevant with regard to the type of database that will be chosen. German, French, Dutch and English were named by most. It was also stated that the database should include questions in all languages of the member states.

3.4 IT Aspects

3.4.1 Technical maintenance

Some stakeholders shared their view on the responsibility for technical maintenance. It was said that technical maintenance should be done by an specialized organization, not by the member states themselves.

3.5 Interests of participants

3.5.1 Should all member states participate?

We asked the stakeholders if it is a condition for a working European database that *all* Member States participate. The responses to this question were:

- “Preferably all, but it is not a prerequisite to start”
- “Don’t force to participate”
- “Wide participation of experts is essential”

This shows participation of all member states in a uniform database is not a prerequisite for starting.

3.5.2 Success criteria

According to the representatives we interviewed, the success criteria of the implementation of an European database are:

- Easy to use / technological accessible;
- All exams must be equivalent and comparable
- The item bank must be constantly updated
- The item bank must include regional content
- Items must be of better quality than the average quality of current items.
- The item bank must guarantee objective exams with multiple choice questions

- Countries must keep (some) autonomy in creating the tests
- Participation in the centrally must not require large investments

3.6 Security

3.6.1 Confidentiality en privacy

The views of stakeholders on confidentiality do not create a clear picture yet. All pointed out that the database must be well secured and that there must be clear procedures with regards to security and confidentiality. For now, it is enough to conclude that member states who want to participate in the uniform database must agree where the responsibilities for security lie: locally or centrally.

As for data protection, stakeholders indicated that EU regulations must be followed. Also, some said that this is merely a national issue. Further discussion about this will be needed later on in the process.

3.7 Support

The need for support in the start-up phase and after implementation was not included in the interviews.

4 Conclusions and recommendations

We begin this chapter with answering the main questions of the survey. Then, we list our recommendations of the steps to take and the action to initiate.

4.1 Main questions and answers

What are the advantages of the development and implementation of a European multiple-choice *database*?

Based on the interviews with stakeholders, as well as on our expertise in the industry, the following advantages were identified.

- Efficiency: developing questions centrally can save time and money, because member states need to develop fewer questions themselves when they make use of questions developed by others. The total workload is less.
- Quality: a centralized database of questions can lead to a higher quality standard of exam questions, because all questions will go through the quality checks that the member states have agreed upon.
- Quantity: a centrally developed database is most likely to contain more questions than the national databases, because the joint capacity to develop questions is larger than that of the individual member.
- Actuality: a common database is easier to maintain and is less likely to contain only outdated questions, because state-of-the-art procedures and supporting software will facilitate scheduled updates.
- Fraud prevention is easier to organize with a central database than with a number of national databases; because participating member states can design procedures together to maximize fraud prevention.
- Analysis and evaluation of items and exams can be organized centrally in addition to or instead of decentrally, with benefits of shared know-how, and opportunities for harmonized difficulty and reliability of the test.

What are the disadvantages of the development and implementation of a European multiple-choice *database?*

- Costs: selection of the software solution, setting it up, making procedures, collecting items and maintaining requires an initial investment in time and money.
- Usage: the time between creation and usage of items may be longer than the current practices in member states, because all questions need to pass the quality checks that the member states have agreed upon.
- Usage: the questions in the database may need to be in a common language for all, because translation and verification procedures are complex and costly in the context of high stakes exams[10]. Verification procedures are necessary costly and will bring on issues with equivalence of the questions. Member states may have additional inconvenience due to need to translate from the source language into their own language of instruction.

[10] Compare, for example, the translation and verification procedures applied in the OECD PISA studies.

- Validity: measuring competencies is difficult with MC questions only, because a task where the right answer must be *recognized* may not be the best way to measure abilities to *apply* knowledge in a given context.

What are the main obligations for participating parties?

Participating member states will need to consider minimal obligations when participating in the central database. The obligations of each member state may differ, because the countries differ substantially in their ability to contribute, as well as in their intended use of the database. For example, it is not realistic to assume that a member state with 10 exam takers per year will contribute equally as a member state with 500 exam takers per year. Therefore, CESNI member states must reach agreement on the conditions for participation, and on the differentiation rules in order to avoid obstacles to participate for some countries.
There is one obligation that needs to be equal for all: each participant need to endorse the quality standards and standardized procedures that the member states (will) have agreed upon. The standards and procedures should comply with best practice in assessment[11].
Participants can contribute in cash or in kind, depending on their 'reasonable share' in the effort. The member states all have experienced experts (some more than others) who are potentially able to develop questions, to review questions, to adjust questions, or to join in a committee that approves items[12]. There is no reason *not* to benefit from this expertise. Contributions in kind by the participants will encourage the support base for the common effort, as well as give an impulse to the quality of the question development procedure. Participating member states may also contribute by submitting previously developed questions from their local database for approval and for sharing with the other member states. Last but not least, participant may contribute by taking part in work groups or project teams aimed at establishing and maintaining the technical, organizational and quality environment.
Each member state is encouraged to nominate a member (assessment expert) for the steering committee. The steering committee is an advisory group that provides strategic oversight and guidance to the project team. Also, each member state is encouraged to nominate (assessment) experts to join a working group. A working group is a group of resource persons who joins expertise in order to work on a specific subtheme or task during the project phase. These working groups can explore and prepare activities such as the contribution of questions from the national databases to the European database, choice of software, quality procedures, guidelines, etcetera.
Member states can choose their own path of participation, e.g.:
participation level 1 = be informed through project updates, provide input for decisionmaking about quality standards and procedures, about conditions for use of the items.
participation level 2 = level 1 + provide expert(s) for question review, for question approval, for sharing item bank / database know-how, for developing procedures, prepare for future use of the questions in the database.
participation level 3 = level 2 + share items, share manuals, share quality assurance procedures, full use of the questions in the database.

---

[11] For further information, the reader is referred to the body of literature on assessment theory and practice. A standard textbook for the industry is Downing, S. M., & Haladyna, T. M. (Eds.). (2006). Handbook of test development. Lawrence Erlbaum Associates Publishers.

[12] Source: interviews with stakeholders. See annex 3.

How much would implementation of a centralized database cost?

The potential costs of implementing a centralized database are hard to estimate. More information is needed to come to an accurate estimation. Factors that influence the cost are the desired time path, the chosen quality standards and procedures, the number of users[13], and the features of IT solution that CESNI is going to decide upon. A state of the art mature item banking system has relatively high costs in terms of annual licenses, but this may be outbalanced by relatively low costs of investments in the start-up phase. The cost estimates below give an indication of the cost range, from a build-your-on-solution to a high-secure, mature test item bank system with full workflow support.[14]

It must be noted here that the estimated annual costs are indicative only, and based on experiences with organizations[15] with similar objectives and approximately 30 licensed users of the software.

Table: Cost scenarios for the first three years.

Scenario	Content quality standards	IT features	estimated annual cost
Basic	Relatively low standards for quality assurance, such as screening procedures and approval. Simple procedures.	Basic build-your-own-solution, e.g. based on SharePoint lists, document l libraries[16] or Office 365 applications	• Manhours; approximately 2.500-3.500 euro per year • IT solution: 1.000-2.000 euro per year
Mature	High quality standards with regard to authoring procedures, question life cycle maintenance, approval.	Third party itembanking system with full authoring workflow support, advanced security guarantees, rich reporting, and compliant with qti industry standards	• Manhours; approximately 5.000-10.000 euro per year • IT solution: 50.000-80.000 euro per year

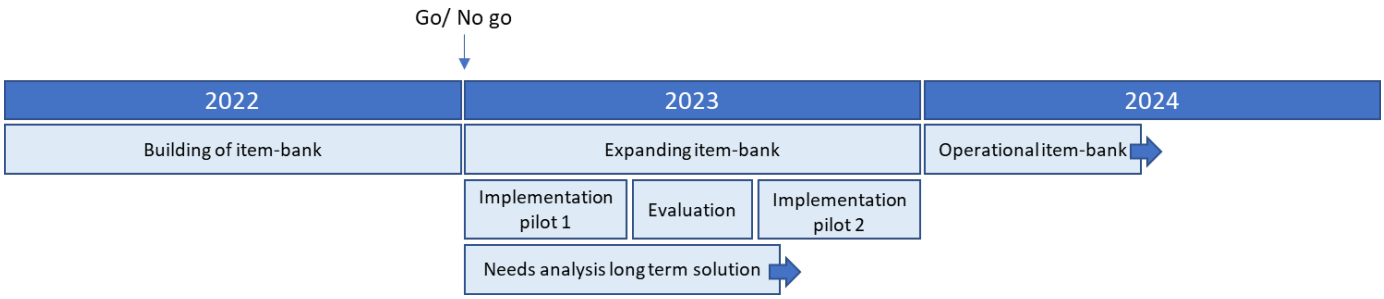
[13] The license costs for commercial software as well as support costs often depend on the number of persons who are allowed to use the software.

[14] The reader is referred to chapter 4 of this report for further details and guidelines for selecting a solution that fits the needs of CESNI.

[15] E.g. Cito, clients of Cito

16 More details on the Sharepoint lists and Sharepoint libraries can be found in Annex 6 of this report, and on the website of Microsoft (Sharepoint List https://support.microsoft.com/en-us/office/introduction-to-lists-0a1c3ace-def0-44af-b225-cfa8d92c52d7; Sharepoint Library https://support.microsoft.com/en-us/office/what-is-a-document-library-3b5976dd-65cf-4c9e-bf5a-713c10ca2872; https://support.microsoft.com/en-us/office/introduction-to-libraries-7d4221d9-8fb9-40d5-8441-2374c84b5e26)

What is the realistic timetable for implementing a centralized database of test questions?
We believe that a timeline of three years is realistic in order for CESNI to implement the common European database. The situation after three years can be described as a fully operational database[17] in place, accessible to all member states who choose to use it. For this to be realized successfully, a number of actions need to be taken and decisions need to be made, which are outlined in more detail in chapter 4. One of the conditions for success is to have a dedicated and skilled project team working on the labour-intensive first three years. In this period, the focus is on coordinating and guiding the team as well as the stakeholders on this journey towards a database setup that can really make member states enthusiastic. After three years, the lessons learned during testing, piloting and 'going live', are likely to spark an interest for future improvements and additional features, for example for the use of non-multiple choice question types. Therefore, CESNI should prepare for decision making after the first three years, in order to choose the path and timing towards a next ambition level.



See Recommendation 0.2 for further details.

4.2 Feasibility and scope of the project

From the interviews, we conclude that there is a broad support for a centralized database which supports examination of candidate boatmasters. There is consensus about the ambition to arrive at a central database so that examination content can be harmonized in line with the ES-QIN requirements for the theoretical exams that will be the new standard as of January 2022.
The interviewees expressed the potential benefits of sharing content with other countries opposed to developing and maintaining exam questions and assessments individually.
The benefits of sharing content are the biggest for countries that have small numbers of candidates. However, countries which already have an organized database, such as CBR in the Netherlands, can also benefit from a uniform centralized European database.

In order to arrive at such a database, some obstacles must be dealt with. One of these is to come to agreement on the scope. Stakeholders vary in their views from "nothing should change for the candidate" to an "all-in-one package" containing linked IT-elements in all phases of the examination process, which does imply change for the candidates.

[17] The quantity of questions at this point, will of course depend on many factors to be decided. For example, we have assumed (see chapter 4) that at least some of the member states will be willing to make questions available from their own database in order to get a good headstart. The conditions to make this possible will of course be subject to decision making and negotiation.

We must realize that currently, there is no existing cooperation of sharing items to build on. Therefore, we strongly advise a step-by-step approach, while keeping the ambition to scale up later.

In this chapter, we narrowing down the scope to a size that is feasible and acceptable for all. Then we list our recommendations and clarify the proposed phasing of the activities in the coming years.

4.2.1 Narrowing the scope

The intended use of a database determines the scope of the solution and has an impact on work processes as well as costs for technology, usage and management (functional and technical).
For each phase of the examination process, one can determine the desired future situation regarding processes and supporting IT infrastructure and make an estimation of the impact on work and costs.

The figure below, shows a ‘high level’ view of three possible scopes using the test cycle as a timeline.

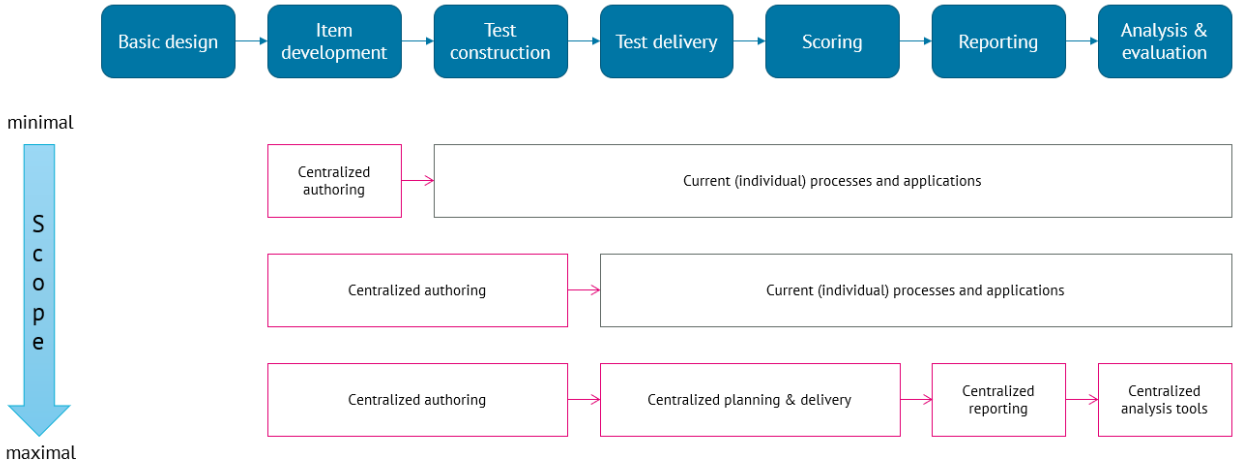


Figure 1 Possible variants considering scope

The arrows between the squares represent an interface: a point where two systems, processes or organizations interact. All variations between minimal and maximal scope are possible, if the interfaces are defined right.
The minimal scope is to centralize authoring of items. All other process are dealt with completely by individual member states. The middle scope is to centralize not only item construction, but also test making. The maximum scope is to centralize authoring, test construction and test delivery.

We advise to choose a solution which allows different stakeholders to enter or leave the centralized parts, in phases. This will make it possible to choose the best solution for each phase of the test cycle.
In Annex 3, a description is given of a mature test platform that combines all the phases in the test cycle.
In the following section, we will describe the three scopes of figure 1 in some more detail.

Minimal scope
The minimal scope consists of a centralized authoring environment: item writers can add questions and those who administer tests can use these as a source for their test construction. This allows the current way of working to co-exist with a new way of working. Participating countries are in control to use the items they want, and add items of their own.
Process: Item authoring, quality assurance and maintenance are centrally organized according to certain quality criteria. All other processes, from test assembly to analysis and evaluation, are the sole responsibility of each of the participating countries.
IT requirements: authoring environment

Medium scope
The medium scope is the situation where harmonization goes as far as the composition of the test papers. A central organization makes sure that the test contains a representative set of questions, that meet the agreed test specifications.
Process: Item authoring, quality assurance, maintenance, and test assembly are centrally organized according to certain quality criteria. All other processes, from test delivery to analysis and evaluation, are the sole responsibility of each of the participating countries.
IT requirements: authoring environment, test assembly tooling

Maximum scope
The maximum scope is the complete chain of centralized solutions for each phase of the test cycle. The whole process is harmonized. For reasons of accessibility, candidates must be facilitated to take the test in their (national) language. If the test is administered online, this could be organized on a anytime-anyplace basis.

Process: All processes, from item authoring to analysis and evaluation, are centrally organized. Translation of items into the languages of the member states, is also centrally organized, as well as research into equivalence of all the different versions.
IT requirements: A (modular) system which supports all processes from item authoring to analysis and evaluation. The systems supports entry, maintenance and presentation of different (equivalent) versions of each unique item in different languages.

4.3 Recommendations

0 SCOPE

Recommendation 0.1 We recommend to keep the scope small in the first stage and build an easy-to-implement database, filled with exam questions from all participating member states.

Clarification:
The main reason to start small is that most stakeholders expressed the importance of keeping test assembly and test administration in national hands. Other arguments to choose for this are:

- Highest chance of smooth implementation
- Low initial costs
- Quick (first) benefits for all member states
- Low risks

The long term solution may or may not be like the solution described as the 'maximum scope'. We deliberately leave this open, because this is most likely subject to future decision making. Once the minimal solution has proven it's *raison d'être* (higher quality standards and cost effectiveness), it can be expected that member states wish to make a next step. Therefore, we advise to not close any doors for scaling up the scope in the future.

What we advise is a leap forward in standardization and professionalization: item content, meta-dating, and storing will be harmonized. Therefore, we define the targeted end product of this (first) step as follows:

- **an item-bank with metadata**
- **a centrally organized process for adding and modifying items**
- **a decentralized organized process for extracting items**

Recommendation 0.2 Make a detailed project planning and take 2 to 3 years for full implementation.

Clarification
The momentum is now, because member states need to adapt to the ES-QIN 2019 competence standards as of 2022.

Suggested timeline
2022

- Establish a small, and preferably, dedicated implementation team which is responsible for building the item bank
- Start with 3-4 countries who are dedicated to the project. We recommend that a base-set is built from items already in use.
- Start with open source or low-cost "off-the-shelf" software tooling with low maintenance costs. Accept that the functionality is limited.
- In the course of 2022, CESNI (steering group) takes a Go /No Go decision to start a pilot with sharing items.
- Start a needs analysis for requirements for a system that can be used in the long term.

2023

- Continue adding items to the item bank
- Evaluate the pilot
- Expand pilot by (1) allowing more (user) countries to extract items (2) inviting other stakeholders to add items
- Evaluate the second pilot
- CESNI/stakeholders set new goals for 2024 and further
- Item bank is fully operational for all member states who wish to join.

2024 and further: consolidate the use of the item bank, make adjustment as the member countries see fit, adjust ambitions as the member states see fit.

1 ORGANISATIONAL ASPECTS

The recommendations in this paragraph relate to in the methodology section (1.4). Where relevant, we have distinguished between the short term goals and the long term ambitions.

Organizational questions*:*

1. Which organization or body should and could become responsible for an international database?
2. Who should be responsible for the content? Who for operation and maintenance?
3. How should this be formally organized?

Recommendation 1.1 Make CESNI responsible for the content, the operation and the maintenance of the database.

Clarification

It is important that the responsibility of the item bank is in the hands of one organization. We consider CESNI suitable for this because CESNI represents the member states, and it appears from the interviews that CESNI has the support from the member states to assume this responsibiliy. . Another option is not logical. CESNI could delegate (parts of) the operational work to a team that is hosted by another organization (in one of the member states). This team should report to CESNI.

Research question*:*

4. How can transition in terms of responsibility for content be organized between the contributing national database(s)?

Recommendation 1.2 Formulate quality standards and procedural conditions for member states to contribute items to the new international database. Create review and approval procedures. Formulate conditions for member states who wish to use items from the item bank.

Clarification

In the current situation, each country is responsible for its own database. Some countries have built a database of considerable size, others have no database at all. It seems reasonable to ask some member states to contribute more than others. This will be necessary to make a good start with a central database of exam questions.

Processes for contribution and extraction of items need to be well-established and agreed on by each participating member. An important aspect of this is the security of items, which should be enforced partly by means of technology (e.g. authentication) and partly through adequate procedures. Leakage of

information to test-takers (in the same or another country) is a risk that must not be underestimated as it can harm the value of the tests and trust in the system.

Research questions*:*

5. How should the work that is related to building and maintaining a uniform centralized database, be practically organized?
6. Who has access to the database / item bank?
7. Item bank ownership; what roles and responsibilities must be defined?
8. Who has access to documentation such as manuals, user guides, instructions, checklists for all members involved?

There are different ways to organize the process of item creation, storage and retrieval. Cito can provide examples of work flows, if needed. We believe it is best to come to agreement on the process in one or two sessions in which all steps are outlined and discussed. Good practices from individual members states can be used and integrated. Access to documentation and to the database is also a topic for discussion and be agreed on.

Recommendation 1.3 Make a (small dedicated) project team responsible for developing procedures and guidelines.

Clarification

Organizing, installing and implementing a database of test questions can be complex. We recommend to assign this task to a project team whose responsibility it is to do so. It should also be the project team who develops quality guidelines, procedures and manuals.

It is important to clearly define who is responsible for decisions with regard to the installation, maintenance and operation of the item bank. This person needs to have sufficient mandate in the organization, and should be able to carry out the task according to rules which are agreed on by CESNI/the member states. The item bank owner can delegate tasks to other team members.

After the item bank is installed and filled with content, the project team can hand over its tasks to others who work under responsibility of CESNI.

The project team is responsible for building the item bank. We recommend to start with a team of at least three people.

- 1 project manager / coordinator of item bank development; assessment expert, organizes item creation; responsible for quality assurance.[18]
- 1 database / item bank manager; responsible for organizing the database; provides user support. [19]
- 1 office manager; supports the workflow of entering data; communicates with ‘suppliers’ such as item developers, reviewers.[20]

[18] Suggested profile: academic, minimum 5 years of experience in leading large-scale item banking projects. Know-how of the main principles of database management systems. Strong communication and negotiation skills.

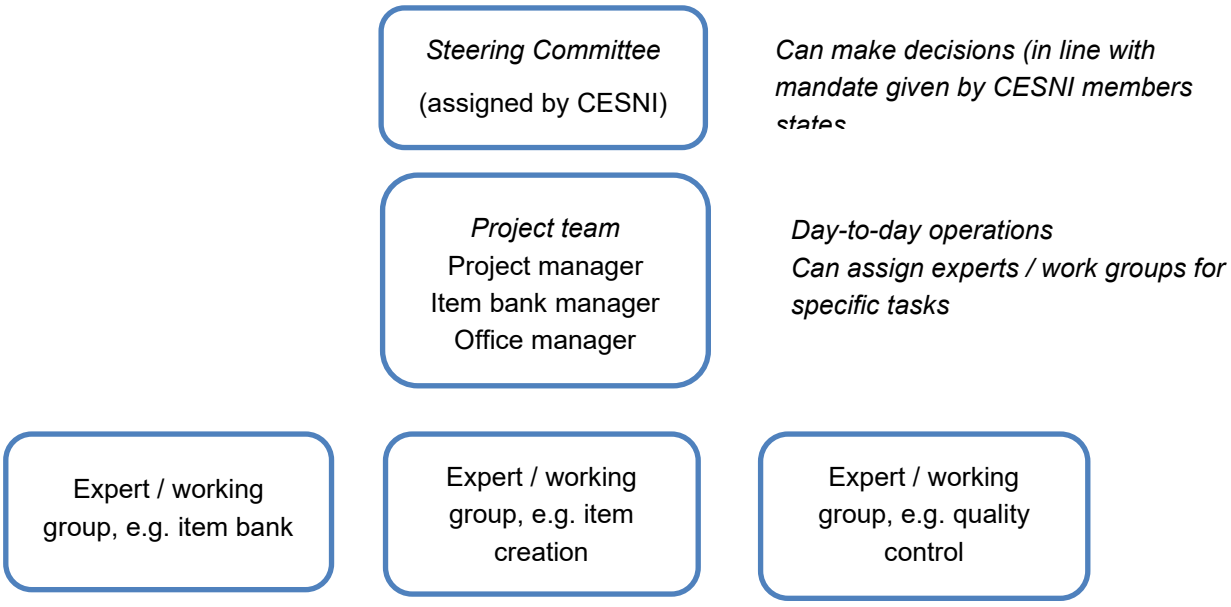
[19] Suggested profile: Bachelor, minimum 3 years of hands-on experience in customizing a database management system. ICT-skills. Able to write instructions, manuals, wiki-pages, user stories.

[20] Suggested profile: vocational diploma. Able to communicate effectively with IT-specialists as well as with users. Organisational skills. IT skills. Problem solving skills. Initiative. Administrative skills.

Tasks of the project manager

- Prepare decision making on the choice of software
- Prepare the structured organization of the metadata in the bank
- Design procedures and workflow
- Describe procedures, roles and responsibilities for users of the item bank.
- Organize item creation, reviewing, quality criteria and acceptance criteria
- Organize and execute pilots with future users
- Temporarily assume the role of item bank owner until this role can be assigned to a permanent staff member.

The project organization can be structured as in the graph blow.



We estimate the required effort for the project team to be 20-40 hours per week. It really depends on the number of items the database should contain and the deadline of the project. We think it is reasonable to assume that it will take 2,5 – 3 years before the database is fully implemented. It can be done in less time, if needed.

Each member state is encouraged to nominate a member (assessment expert) for the steering committee. The steering committee is an advisory group that provides strategic oversight and guidance to the project team. Also, each member state is encouraged to nominate (assessment) experts to join a working group. A working group is a group of resource persons who joins expertise in order to work on a specific subtheme or task during the project phase. These working groups can explore and prepare activities such as the contribution of questions from the national databases to the European database, choice of software, quality procedures, guidelines, etcetera.

At this stage, it is hard to estimate how much time is involved in participation in the steering committee and in working groups.

At an operational level, the use of a central database comes with the responsibility of managing functional and technical issues that will arise in working with the database. A plan should be made for embedding

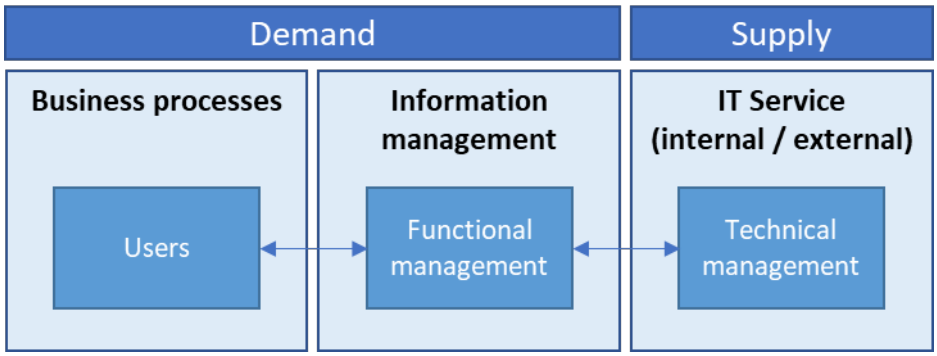
this in the organization. A typical way of organizing this is by implementing processes and roles for functional management and technical management.

The areas of focus are roughly:

Functional Management	Technical Management
- Support of the end user, e.g.: functional questions, training, incident and authorization management. - Linking pin with Technical management	- Responsible for application configuration and maintenance, server and database configuration and data backup. - Linking pin with IT infrastructure

The size and organization of these processes depend on the scale of the chosen solution but aspects of both roles will always have to be addressed to a certain extent. Within the small scope we recommend, this could both be handled by the same person.

Functional management is necessary on a “day to day”-basis to support users and because of the necessary knowledge of the user domain it is recommended to make this part of the same organization that is responsible for the centralized database. Technical management, if necessary, can be the responsibility of the solution provider (internal/external IT service).



2 TECHNICAL ASPECTS

Research question:

1. To what extent should the database consolidate all aspects of exam management into one system with a single view, enabling access, control, sort, analyze and export items into an exam delivery system?

Recommendation 2.1 Ensure that the database enables import, export and publication of items. These functions will do for the first phase.

Clarification
Currently, members states administer either oral exams, paper and pencil exams, or digital exams. In order to facilitate exchange with the national organizations, the supporting software needs to contain the following import and export functions:

- Import of test items and metadata[21] through templates/forms
- Export of items in various formats: Word processor and QTI (IMS Question and Test Interoperability specification);
- View items on screen for editing purposes
- Link with existing examination environments.

General principles

Mature item banking systems use industry standards for data exchange, i.e. “IMS QTI” for exchange of test information and “IMS Results reporting” for exchange of assessment information (candidate results).

By choosing for a system that uses these industry standards, CESNI will be well prepared for future growth towards support of the full chain of construction up to reporting. Also, using industry standards will facilitate exchange with member states who already have a mature item banking system in place.

Research question*:*

2. To what extent should the database provide a centrally managed repository for items?
3. To what extent should the database provide an exam delivery system for candidates
4. To what extent should the database provide custom data analytics and reports?

Recommendation 2.2 The database should contain basic functions, such as select and sort. These functions are needed for efficient retrieval of questions.

Clarification

The basic functions such as select and sort are common item bank functions that should be available to end users. Based on the interviews, there is no urgent need for functions such as exam delivery, custom data analytics and reporting. CESNI could choose to organize a detailed needs analysis in order to point out what other database functions may be needed in the short run. A detailed needs analysis was not included in the current study.

Taking it a step further than a database only, we list the long-term advantages of developing a complete exam system, which includes item creation, test creation, test administration and analysis. These considerations might get back on the agenda after the short term goals have been achieved.

Uniformity: Using a harmonized way of examining candidates will ensure equal chances for all candidates. The quality of examination will not depend on the quality of examiners. Comparison between exams will be possible when both the exam as well as the process of exam taking are the same for all candidates.

Efficiency: Developing tests centrally will be cost and time effective for participating member states. It will also reduce the handling time per candidate.

Reduction of IT systems and/or processes: A central exam system enables phasing out current IT solutions which will have a positive effect on operational and maintenance costs. A uniform process, supported by uniform IT solutions is more cost efficient.

A complete exam system, which includes item creations, test creation, test administration and analysis has not only advantages. Below, we list the most important obstacles.

Time and money: Setting up an exam system requires considerable investments, both in terms of software fees as in staffing needed for technical and content maintenance.

[21] An example of a simple structure for items and metadata is derived from the ADN catalogue of questions, and presented in Annex 5.

Less flexible: A centrally organized exam system, including the organization to operate this, may lead to a slower response on required adjustments. Also, if such a system rules out oral examinations, individual customization will be lost and tailoring exams to regional needs may be more difficult. Seen from a technical perspective, it doesn't allow for choosing the best fitting solutions for supporting the test cycle and poses the risk of vendor lock-in.

3 CONTENT ASPECTS

The interviews showed the need for a solid content creation and management process. As content domains are not stable over time, due to changes in regulations, items must be checked regularly. Good metadata (to make filtering by content category and checking items possible) must be applied.

Content-related question*:*

1. What should the database consist of? Is it merely a collection of questions and answers? Or also software to prepare an examination, choosing from the pool of questions the set of questions put to a candidate?
2. How should the database be set up and maintained regarding content? (Item life cycle; re-use of items).

Recommendation 3.1 Assign a project manager who is responsible for the definition of metadata which characterizes item life cycle and re-use.

Clarification:
Based on best practice in assessment, we recommend that decisions on how to structure the content, the metadata and the life cycle maintenance should be left to an item bank expert. By doing so, CESNI can benefit from best-practice experiences in item bank maintenance.

The item bank should be able to support all stages in the life cycle of an item:

- construction and review phase
- ready for use (once or many times)
- revision phase
- retired and/or made public

The lifecycle status should be considered as a specific user defined attribute (metadata). The best way to register this information dep30ends on business rules that define how to handle each of these stages. On related note, it could also be useful to register when and where the item has been used, for how many candidates etcetera.

Research question*:*

3. What could be done to make sure that pedagogic / educational expertise is applied when elaborating and evaluating the questions and answers for the database?
4. How can the quality be ensured? In terms of quality of content, quality of translations, psychometric quality.

Recommendation 3.2 Accept only items that have passed a thorough quality review procedure.

Clarification:

Joint quality assurance is one of the main factors that would make members enthusiastic to use the item bank. Only questions that meet common quality standards for educational assessment should be approved for the item bank. A review and acceptance procedure needs to be developed in line with best practice in educational assessment. In the short term, it is most feasible to focus on quality measures to be applied in the construction phase. In the long term, it can be explored how item quality can be brought to a higher level by means of data analysis, to be performed by each member state, or centrally organized. Also, data analysis can be used to check for equivalence of the translations.

In the long term, the option to support the entire editorial workflow functionally by a dedicated software system may be considered, as such mature systems exist on the market and can make processes run better and more (cost-)efficient.

Research question*:*

5. In how many languages (and which) should the exams be available?

Recommendation 3.3 Collect, maintain and provide the items in one language only, and we suggest that language to be English or German.

Clarification:

Based on best practice in item development, we strongly advise not to aim at having (supposedly) equivalent versions of questions in different languages, at least not during the construction phase. It would make things very complicated and costly to allow for different language versions in the phase that frequent reviews must lead to an approved version. Also, this would slow down the process and cause validity issues. Translations creates subtle but often notable differences. Therefore, it is important that there is one 'true' version of the question. After approval, the question can be translated into the language of instruction of the member states, where the user has to be aware that the translated question may not be fully equivalent to the one in the 'true' version in the source language.

Our preferred option for the source language would be English for a number of reasons. Firstly, most of the respondents to the interviews have indicated English to be one of the languages that should be supported. Secondly, it is highly desirable that experts involved in construction and review have at least a working knowledge of the source language and it is our impression that such experts can more easily be found if English is chosen as the source language[22]. Thirdly, a choice for English language does not favour any of the member states, hence all member states have access on equal grounds.

There are good reasons to choose German, however. More discussion is needed to come to a solid decision on this. We recommend to make choosing the language part of the initial phase of the project.

With regard to translation *after* approval, CESNI can of course make other choices depending on the budget. We mention in this respect that the stakeholders from the CCR (Rhine commission) countries unanimously mentioned the importance to have at least versions in German, French, English and Dutch. The stakeholders from the Danube commission and Sava commission expressed a wish to have translation into the languages of all member states. We recommend that test-takers take the test in their own language of instructions, because any other language would put them at a disadvantage. The translation can be organized either centralized or decentralized. We recommend a decentral approach in the short run, which would mean that each country is responsible for translation for their own use. This

[22] One could argue that German is also candidate for being the source language, because it is the official second language of communication on the river Rhine. On the other hand, mastery of German for the purpose of technical communication on the river is not the same as mastery of German for question development or review.

recommendation is in line with the autonomous responsibility of each member states for examination. Also, such a choice would help the project team that is responsible for the start-up, to keep the scope of the project manageable.

In the long run, the database system could grow into a multi-language system, with centrally coordinated translation. If CESNI chooses to do so, we strongly advise to apply sound verification procedures[23] , in order to guarantee equivalence of the different versions.

CESNI should decide on central or decentral translation while taking into account the agreed quality standard of the questions in the database, as well as the interests of the member states.

Research question*:*

6. Should there be any other item types, than MC?

Recommendation 3.4 Choose for multiple choice questions only in the short term, keeping options open for other item types in the long term.

Clarification:

The need for a variety of items types varies between the stakeholders. Most stakeholders will be satisfied with multiple choice questions only. Some stakeholders (e.g. CBR) have expressed a wish for other item types, such as fill-in-the-gaps and graphic interactions. Also, use of video and images could facilitate richer context, which is a potential advantage when the assessment objective is not only knowledge, but skills as well. The respondents mentioned advantages of technology enhanced questions (e.g. simulations, interactive types) in order to assess competences. Also, constructed response (open-ended) and drawing have been mentioned. These options are worth considering, but require in-depth needs analysis and substantial additional investments.

For practical and financial reasons we advise to begin with MC questions only, and prepare for including other item types at a later stage.

4 IT-aspects

Research question*:*

1. How should the database be set up and maintained, with regard to IT?
2. What logical options are there for setting up a database that serves both short term goals and long term goals?

- **Recommendation 4.1** Conduct a brief study for the selection of the software. Consult member states who have experience with this. Make a decision after comparing the pros and cons of available options.

Clarification:

[23] Compare, for example, the translation and verification procedures applied in the OECD PISA studies.

It is beyond the scope of this advice to provide an overview of all options and features. Member states the Netherlands and Belgium have experience with a mature item banking system. At this stage, we limit ourselves to presenting three different directions for a solution, with examples of software that fit in each of the directions. In the short term, a simple solution with basic functionality may be good enough to start building up the content.
Questions which play a role in the decision-making are:

- How important is the exchange of files with Word processor and QTI format?
- How clear is the long term goal?
- To what extent should CESNI prepare for accessibility for test-takers with special needs? (WCAG-compliant)
- How tight is the budget?
- How much capacity (database expertise, manpower) can CESNI arrange to set up a system that is not yet ready for use?
- What is CESNI's policy on cloud based solutions?
- How much capacity is required? (number of items, number of users, video)
- How fast do you need the software?

Although at this stage it is difficult to provide exact costs that are related to the IT-environment it is possible to give an estimation. For this we consider three solutions. They are arranged from least 'ready for use' to most 'ready for use':

Software	Features	Vendor support	Word/QTI	estimated annual cost
SharePoint Plan 1	A basic collaboration environment. The database can be realized by SharePoint lists, document libraries or using Office 365 software	basic	Word/Office	€600,- *10 users. without Office licenses*
Tao Ignite	an online digital (only) assessment platform with features ranging from authoring to reporting. Out-of-the-box collaboration and workflow support is basic.	basic, more at additional cost	QTI, standard offer has no support for export to office documents. Possibly at additional costs.	€12.000,-
GradeMaker Pro	an online assessment platform but capable of also delivering paper based exams and with extensive features for security, quality assurance and workflow.	included	Word and QTI	€50.000,-

It must be noted here that the estimated annual costs are indicative only. Prices can vary because they are subject to volumes, additional features or customized plans (e.g. authoring only) and negotiation.

It is worth mentioning that TAO is also available as open source solution. Although there are no costs related to the software, it is not necessarily a cheaper option. Costs of TAO software consist of hiring IT-staff for customization, infrastructure (e.g. hosting) and maintenance. These are services that are integrated in license tiers or can be provided by the vendor or integrations partners.

A similar comment is applicable for the SharePoint solution: SharePoint (as well as any other database-like solution) is flexible, but not ready for use. It requires effort to customize such an environment for the purpose of building an item bank. Therefore, the solution is not necessarily cheaper but still worth mentioning because of some other advantages. A notable disadvantage of a SharePoint-solution is that it does not support QTI.

Research question:

3. What kind of hardware is needed to appropriately operate the software?

Recommendation 4.2 Decide whether data may or may not be stored in the cloud.

Clarification:

Software vendors can provide detailed information about the system requirements for smooth operation of the software, including hardware requirements. Cloud-based solutions require less hardware for storage, and may be cheaper. The pros and cons of cloud-based solutions should be thoroughly evaluated by CESNI, and a decision taken before choosing a software solution because the policy may limit the options. Furthermore, hardware requirements can be compared based on vendor product information.

5 Interests of participants

Research questions:

1. What could make countries enthusiastic about or what could discourage them participating in a European database?
2. Will the central database provide added value to Member States, for example if they can focus resources on content of questions instead of carrying out [supervising] examinations?

Recommendation 5.1 Make sure that each member can benefit from the new database, it should not be a step back for any of the member states.

Clarification:

This basic rule should be a guideline in order to make member states enthusiastic about participating. Based on the interviews, we conclude that potential benefits are not equally shared among the member states. Members that have a full up-to-date database and mature test platform in place, apparently have little to gain and much to offer. Given that participation is voluntarily (we found no indication that participation may be forced), each member state must weigh its own costs and benefits. CESNI needs to take these potential differences in interest into account when making agreements with individual members on participation in the database. Together, the member states have the capacity to make this work and CESNI can succeed if it can share the benefits and the costs in a way that it is attractive for each to step in.

The main benefits are:

- better test quality (for all)
- test items available in English language (for some, e.g. CBR)
- more capacity to develop items (for all)
- efficiency
- easy exchange (for all)
- more choice / full item bank (for all)

The main obstacles are:

- risk of ‘double investment for those ‘states that already invested in question renewal for IS-QIN19 (some)
- lack of money (all)
- slower question development (all)
- tailoring exams to regional needs may be more difficult (some)
- loss of autonomy (possibly in the long run)
- translation (some/all)

6 Security

Research question:

1. How can confidentiality be ensured?

Recommendation 6.1 Choose for a solid and reliable authentication procedure for all users.

Clarification:

The procedures and technical security measures need to be fine-tuned to the actual risk of leakages.

General principles

The more people can see the items, the higher the risk of unwanted leakage of information. The higher the commitment of people involved, the lower the risk of leakage. Built in a mature system is the feature that users see only what they need to see and nothing more. The item bank owner is in control of assigning user rights and admission procedures. In this respect, an audit trail could also be of value in order to track information on access, modifications, and export. The more ‘primitive’ the chosen software, the higher the organizational efforts need to be to assure the minimally required level of security.

Research questions:

2. What privacy matters need consideration?

Recommendation 6.2 Use procedures for data registration that are compliant with GDPR (EU-General Data Protection Regulation)

Clarification:

At any point where personal data are stored, checks are required whether it complies with the relevant laws. The first time to do this check, is when the procedure to register data is designed.

General principles

Storing user data for the purpose of an audit trail may be justified as ‘specified, explicit and legitimate purposes’ under GDPR. If - in the long run - stakeholders decide to join hands for centralized test administration, the GDPR will also apply to personal data of candidates. We assume that the member states have the necessary know-how on this matter, because they are already familiar with handling personal data of candidates in the national context.

7 Support

Research questions:

1. Which contribution can be expected in the start-up phase from organizations such as CBR?

Recommendation 7.1 Formally invite CBR and up to three other ‘frontrunner’ organizations to share (A) part of their item bank, and (B) their know-how, procedures and manuals on the authoring and item banking process.

Clarification:

CESNI must not invent the wheel when the know-how is already among its member states. The conditions for sharing should be thoroughly worked out and agreed on. The interest of these organizations that are willing to share should be well-protected, and the benefits made clear.
The frontrunner organizations together could contribute a ‘starter set’ consisting of items that cover the whole of the ES-QIN boatmaster learning goals. The competences cover approximately 400 knowledge and skill descriptions, therefore a starter set of 400 questions (one for each skill/knowledge) would make a meaningful start. If the frontrunner countries are willing and able to share more, this is of course even better.
A group of four frontrunner organizations may be considered a sufficiently large ‘critical mass’. The project will then have a solid foundation to build upon and to grow.

Research questions:

2. What support is needed, or worth considering for sustainable database management after delivery?

Recommendation 7.2 Consider involving external support during start-up of the item bank.

Clarification:

During start up, additional support can be useful to speed up the building of the item bank, and to make optimal use of best practice. We mention a few examples of activities where support can be beneficial:

- needs analysis/ requirement analysis for software and hardware
- inventory of useful docs with the member organizations, manuals etc - so as not to invent the wheel
- inventory and first scan of available items among stakeholders
- set up review process
- review the starter set of items

Cito can provide most of these services, should CCNR / CESNI choose so.

Getting started

The upcoming phase of this project will be one of further decision making and aligning interests. The process of coming to this report has been an early and necessary step in that phase. It is now up to the

stakeholders to keep the momentum going. The goal of having an elementary item bank in place within 2,5 – 3 years can be realistic.

We therefore wish to encourage CESNI to:

- endorse the recommendations in this report
- take decisions on several points indicated in this report
- assign a project team to start building the item bank.

Annex 1: Interview questions for stakeholders

Interviews CCNR – set up
Semi structured interview
1,5 hrs in teams
INTRODUCTION (5-10 MIN)
• Background / reason of the research
• Reason for the interview: gathering information of stakeholders
• Set up of the research; *not* a feasibility study; *not* a study for requirements; but it is *to support decision making about a possible database project*.
• Set up of the interview: exchanging thoughts / questions / concerns
○ current situation of exam making, exam taking and reporting
○ future situation; working with a centralized database
Consent for recording of the interview. **Yes / No**
Name
Function / role
Date
CURRENT SITUATION OF EXAMINATION (20-30 MINUTES)
1) How does examination take place now and in what form?
a) Process of item (question) creation
b) Process of test making / reviewing / improving
c) Process of test taking; the way candidates take exams
d) Process of reporting on results
e) Which technical aspects should be taken into account in the development, maintenance and administration of the exams?
2. Pros and cons of current practice

USE OF A CENTRALIZED DATABASE / TEST PLATFORM (30-40 MIN)
General
3. What are your thoughts on working with a centralized database or platform for exam questions?
a) What would it imply for your organization in terms of organization, technology and finance?
b) What would it imply for your exam takers?
c) General advantages
d) General obstacles
4. With regard to a centralized database of questions, what is important for you and your organization? What would be leading principles when developing a centralized database? (Quality, autonomy, cooperation, efficiency, clear responsibilities, future proof, software, financial … …)
5. In your opinion, when should a European database be available to the different Member States?
6. Is it a condition for a working European database that all Member States participate or is that up to a Member State?
7. What should such a database do most? What could make you / countries enthusiastic about or what could discourage them participating in a European database?
8. Is it possible for you to contribute to the content or maintenance of a European database?
9. If so, what kind of contribution are you thinking of? (Knowledge (e.g., drafting questions); financial; staffing?)
Content aspects
10. What item types; only MC? Or more in future?
11. Specific questions for own context? How many in total?
12. What language; now and in future?
13. What content domains?
14. Would illustrations / video be of added value for the quality of the exam?
15. Would you like data to improve questions for future test takers?
16. Versions of exams / renewal of items/exams?
Test taking
17. Where do tests take place? Now and possibly in future?

18. When do tests take place? Set moments or flexible?
19. How many exam takers per year?
Organizational aspects
20. Who should (always) be responsible for content and maintenance of content?
21. Who should (always) be responsible for the functioning of the database?
Technical aspects
22. Who should be responsible for technical maintenance?
Security
23. How can confidentiality be ensured?
24. What privacy matters need consideration?
Final remarks
25. Do you have anything to add?
26. Do you have questions / issues …?
NEXT STEPS – AFTER THE INTERVIEWS (5 MIN)

Annex 2: Schedules interviews with stakeholders

Germany	Ms Petra Nethövel-Kathstede	petra.nethoevel-kathstede@bmvi.bund.de	17 Aug
Netherlands	Ms Jolanda Kliest, Dutch Ministry Ms Katja van den Beld, CBR	jolanda.klies@minienw.nl katja.van.den.beld@cbr.nl	2 Sep
Switzerland	Ms Loredana Foselli Mr Lukas Sibler	loredana.foselli@portof.ch lukas.sibler@portof.ch	23 Aug
Belgium	Ms Herlinde Liégeois, De Vlaamse Waterweg	herlinde.liegeois@vlaamsewaterweg.be	2 Sep
Belgium	Ms Di Luzio Service public de Wallonie	aurelie.diluzio@spw.wallonie.be	6 sep
Danube Commission	Mr Igor Alexander	igor.alexander@danubecommission.org	17 Sep
Sava Commission	Mr Zeljko Milkovic	zmilkovic@savacommission.org	20 Aug
Luxembourg	Mr. Max Niles	max.nilles@tr.etat.lu	15 Okt

Annex 3: Summary Report of interviews

Current situation of examination

1 How does examination take place now and in what form?

a) Process of item (question) creation

- catalogue of open questions that can be asked verbally or in writing, from which items are selected; examiners conduct an oral examination, which may or may not be combined with a written examination; new questions are needed; Pros: Oral is flexible; Individual customisation is possible; Misunderstandings can be resolved so that the candidate still understands the question correctly; Con's: Current catalogue is out-dated; Quality of examination (reliability and validity) depends very much on the quality of the examiners; Quality of exams differs enormously now; Comparison between exams virtually impossible;
- catalogue of open questions that can be asked verbally or in writing, items are chosen from this catalogue, no constructions are made, the choice is made manually by the examiners. ExCie prepares the questions and makes overviews of cards to be used and assignments to be carried out by the candidates.
- done by the examination committee, which is divided into subcommittees. Theory exam for skipper consists of quite a lot of questions, in some parts 4 times the length of the exam. Current questions are outdated, huge job to keep it up to date.
- On the basis of the test matrix, CBR has multiple-choice questions with three alternatives developed exclusively by free-lance item authors. They write questions and answers, and CBR assesses and screens the developed questions for technical and linguistic issues. The contents of the questionnaires are then assessed by a technical committee

b) Process of test making / reviewing / improving

- no construction anymore; no changes;
- no fixed list of questions or something of a database questions. The examiners, from the experience they have and under their own control, use questions from the previous exams.
- No reviews, but sometimes adjustments to the questions of the catalogue due to new rules. There are 10 experts in the Exam board
- By Exam board
- CBR prepares test versions of the approved questions on the basis of the test matrix and submits these test versions for approval to a Board of Experts with 5 experts by experience
- Candidates register themselves for a digital exam, at a time and location of their choice. The digital exams are taken at CBR locations throughout the country. Contains hot spot and numerical questions

c) Process of test taking; the way candidates take exams

- Oral examination, combined with a written examination. 5 to 7 locations in Germany,

- test in writing. Then the examiner is doing the short overview of the result. They show you were you are wrong. Only 2 marks, pass or not pass. After written exam, after you pass you can go to the practical exam.
- Written examinations with open questions and card assignments; The examinations have been outsourced to an external party (Cintra), the examination system has been taken over. The theory exams are administered digitally.

d) Process of reporting on results

- The examiner communicates the results.
- An exam protocol is filled in with score points, immediately after taking and assessing the answers, on the same day. Candidates will only receive a pass/fail verbally.
- Candidates receive the result immediately after the test: failed or passed. If they fail three modules, this is reported and they only have to retake the examination for these modules.
- Results to candidates via a result form, linked to the final attainment levels. Results are divided according to the end terms. Discussion of results and pass rates in technical committee and BoE (with Ministry), possibly broken down by trainer.

e) Which technical aspects should be taken into account in the development, maintenance and administration of the exams?

- New questions are needed, regular updates are needed due to new laws and regulations, and quality monitoring is needed: especially the difficulty level of the MC questions
- first of all basic simulators, then other hardware or software will be very helpful
- What software is needed to develop questions?; Login to the database required; What will be the process of developing and reviewing/adapting?; How is the reporting done? referably 'all-in-one-package
- The collection system is cumbersome and outdated.
- Discussion of examination requirements in technical committee; Psychometric analysis of all test variants

2. Pros and cons of current practice

- Pros:

  - Oral is flexible; Questioning is possible; Individual customisation is possible; Misunderstandings can be resolved so that the candidate still understands the question correctly

  Cons:

  - Current catalogue is out-dated; Quality of examination (reliability and validity) depends very much on the quality of the examiners; Quality of exams differs enormously now; Comparison between exams virtually impossible

- System is old. The system must be modernized. Schools have to adjust the programs. Also navigation time. Not enough time on ship. There are no school ships like I saw in the Netherlands.

- Pros: - Now it is good for the candidate: quick result known; Now everything is in own hands for composing questions, customisation is possible; Now 8 to 9 candidates can be examined per day; Now different exams for different routes for different candidates at the same time possible

  Cons: - Preparation/selection of questions and cards takes a lot of time (approx. 30 min per candidate); Secretariat spends approx. 1 hour per candidate on administration before/after; Many experts needed per test day (about 3-4); Examinations take half to full day

- Pro: digital, immediate results

  Con: log system, limited number of questions which increases the risk of leaks (although there are no indications of this), no practice exams.

- Pro: Item bank with closed questions, is fraud-proof, and good analysis of exams and assignments is possible. More eyes view exam. Unique assignments in exam.

  Manageable examination system, through well-established procedure. Valid and reliable exams.

  Cons: Absolute right or wrong answers to multiple choice questions.

  It is a linguistic exam, aiming for B1-level, without using jargon. Extra time needed for less language competent candidates.

  Exams now only in Dutch

Use of a centralized database / test platform

GENERAL

3. What are your thoughts on working with a centralized database or platform for exam questions?

- Topicality is important; - Question is which program (IT) will be used for development and purchase; No financial insight either; Amount of work for preparations by Exam Committee is important
- Questions would be developed and translated centrally, input from the central European body. It is important to agree on the development, approval and adoption procedure. This also applies to the composition of the versions. Tailoring to regional situations is more difficult: CBR focuses on regulations and NL inland waterways. Psychometric analyses still possible. No more management of bank and statements - slower responses, because adjustments to statements are directed centrally.

a) What would it imply for your organization in terms of organization, technology and finance?

- Easier for them to do this financially. It will not help so much financially, though. (2). They will now be clear that they have the database, the burden of the work will be less, more quality. Then they can rely on the database. I am sure that the database will be welcomed certainly. But the contribution to the database and the questions must be wide from the whole network.
- None
- More attention to the components that are not generic, for which more items can be developed than is currently the case, more attention to quality.

- Candidates should not notice this, but regional differences in questions should remain.

b) What would it imply for your exam takers?

- Possibly lower costs for the candidates in the longer term; The textual nature of many questions may make it more difficult for some candidates; Language can become an issue for some candidates; No extra explanation possible as with the oral examinations); General advantages ;
- Yes, they will be also thrilled if the database will be available for them

c) General advantages

- it will ease of process for examiners; it will form unique or harmonized way to exam the candidates; it will help the candidates; for state administration it will be great help from the point of simplicity to rely on such database
- It is important that the same questions are available everywhere: harmonization is very important (because of the EU rules)!
- Higher quality; more uniformity
- All exams are based on the same guideline (content-wise all the same and within Europe everything is harmonized) and uniformly administered (digital multiple-choice),
- No more differences in exams, all exams equal, no more written exam with open questions

d) General obstacles

- may require a lot of preparation (organizational, task distribution, financial aspect, monitoring and revisions of the database
- none
- Database maintenance: who and how?; One central organization is needed for the management of the database; Security (secrecy) can be an issue
- The Dutch inland shipping sector would like to have exams in English, because of the many foreign lateral entrants to NL. However, CCNR regulations state that boatmasters must speak Dutch or German.

4. With regard to a centralized database of questions, what is important for you and your organization? What would be leading principles when developing a centralized database?

- It is important to guarantee that the database is constantly kept up to date and is therefore immediately adapted in the event of changes to the law or new rules. Harmonization is also important: everyone should use it.
- security; One central administrator; Preferably a web-based solution
- It must be clear who is responsible. Many experts are now working on it, specialists, that must be preserved. Validation is important. Well organized secretariat necessary.
- National and regional differences must be preserved. The countries' own influence and contribution are important in the case of a central database and must be properly regulated. In case of problems in questions, fast reaction and solution, also in case of central database. NL wants to have influence on the number of exercises and the distribution in the test matrix. NL wants to have influence on the inclusion of regional differences in exams.
- In other words, a say in the quality of the exercises and the exams.

5. In your opinion, when should a European database be available to the different Member States?

- By January 2022 preferably, but a little later and better than quick & dirty
- As soon as possible!
- End of 22
- Make realistic time path
- As soon as possible
- 5 years

6. Is it a condition for a working European database that all Member States participate or is that up to a Member State?

- Preferably, but it is not a prerequisite to start
- You cannot force anybody to participate. Wide participation of experts in forming the database will be essential. From the very wide base of experts in Europe, and then we will have a good product. We must ensure the very wide participation in the formation and the maintenance of the database.
- Preferably all of them, but definitely Germany, France, Belgium, the Netherlands and Switzerland
- All should decide for themselves
- From the point of view of harmonization, all countries should participate. But participation depends on the number of participating states and whether, for example, neighbouring countries are involved (in our case, Germany and Belgium). It is important to have a spread of participating countries. A database of all 5 CCNR countries is already a great step forward.

7. What should such a database do most? What could make you / countries enthusiastic about or what could discourage them participating in a European database?

- Must be easy to use and constantly updated.
- More languages for candidates
- less handling for exam organizations
- technology accessible;
- content should cover your part of the world
- Deliver higher quality items
- Guarantee quality of exams with multiple choice questions and that all exams are equivalent and comparable. Saves time and money for other countries, because development of assignments is centralized. Database must be easy to use, connection must not require huge costs. Countries must feel that they retain some autonomy. So countries must be able to increase the quality of their exams at low cost and effort. This applies in particular to countries with a limited number of inland navigation examinations.

8. Is it possible for you to contribute to the content or maintenance of a European database?

- Experts from the ministry department responsible for examinations may contribute.
- Yes

- Current experts can help develop, review, adjust, etc.
- Yes
- contribute to the content and maintenance of a central database.

9. If so, what kind of contribution are you thinking of? (Knowledge (e.g., drafting questions); financial; staffing?)

- Delivering content, developing items, reviewing developed items
- Financing is big minus
- 8 to 10 experts
- Bringing in knowledge about the content of exercises and the psychometric functioning of exercises.
- translation

CONTENT ASPECTS

10. What item types; only MC? Or more in future?

- MC, multiple response, interactive questions, simulations
- MC on a level, not only text but also illustrations (photos, drawings, and possibly video/audio), certainly also more competence-based (i.e. not only aimed at reproduction/knowledge but also at application, i.e. more skills-oriented)
- Also other formats may be possible such as using cards (drag&drop); Video tutorials; Open questions; Drawing on the computer
- MC
- Mc with 3 answers, due to difficulty to develop 4th alternative. In the future, variations on closed questions, such as hotspot, yes/no, fill-in-the-blank, sequence question

11. Specific questions for own context? How many in total?

- No
- Mainly general, EU rules; now the number varies per exam (police-rules: about 50-60, further per topic about 20 questions available, in new database about double number needed
- Now German, soon French and Dutch and English may also be possible. For countries like Slovenia, Romania, Poland and Hungary (which now take exams in Switzerland), German is definitely needed.
- 10%
- 20%

12. What language; now and in future?

- German, Czech, French, Russian, English, Dutch, Ukrainian
- German, French, Dutch and English.
- Language of the Member State
- German, French, Dutch and English

- Dutch and English
- preferably all languages used in member states, but this can be a problem
- If we have at least one language, then each country can do the translation for themselves
- This is a matter of financing

13. What content domains?

- The EU legislation
- Certainly content from the CESNI tables.
- Police and maps; The police rules and European rules of the different Member States
- According to the new content domains
- All theoretical examinations on competences for the inland skipper.
- knowledge of local rivers

14. Would illustrations / video be of added value for the quality of the exam?

- yes
- Yes!
- Certainly
- Yes
- Photos': yes; video: no
- yes

15. Would you like data to improve questions for future test takers?

- Yes
- Yes
- Yes
- Yes
- yes

16. Versions of exams / renewal of items/exams?

- Make sure it is not easy for students to learn by heart and pass exam
- Up to 4 versions should be possible for taking up to 4 candidates at the same time
- Yes, dynamic of course. Not useful to have one fixed kind of exam.
- 4-6 versions
- Preferably one large item bank
- Yes, CBR wants to continue to compile its own exam versions and make its own psychometric analyses.

TEST TAKING

17. Where do tests take place? Now and possibly in future?

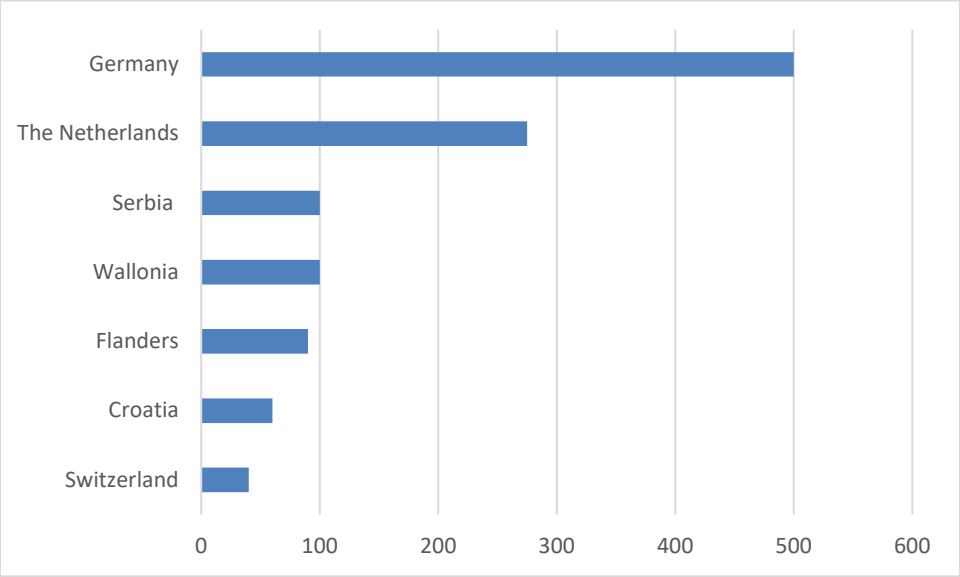
- Now at 5 to 7 locations in Germany, keep it that way at first, but may be able to limit it with centralized test taking
- 2 locations

- 1 location
- Various locations

18. When do tests take place? Set moments or flexible?

- depends on the availability of the examiners (who do not have this work as their main job), there are no 'fixed dates' now.
- Fixed dates
- Twice a week
- Flexible moments

19. How many exam takers per year?



Note: the numbers are educated guesses only, not to be cited without solid verification.

ORGANIZATIONAL ASPECTS

20. Who should (always) be responsible for content and maintenance of content?

- Consulting firm
- EU commission
- Preferably a neutral body, i.e. a third party, not a stakeholder
- It must be certainly one strong and rich organization or entity
- Cooperation of the examination boards
- CESNI
- An independent body should carry out development, maintenance and management of item bank.

21. Who should (always) be responsible for the functioning of the database?

- Preferably a neutral body, i.e. a third party, not a stakeholder
- EDINNA (the educational network of inland waterway navigation schools and training institutes)

- In principle, one central organization, but if the exams become web-based, there is accountability per Member State: in Switzerland, they are very flexible.
- An independent body should carry out development, maintenance and management of item bank.

TECHNICAL ASPECTS

22. Who should be responsible for technical maintenance?

- Preferably a neutral body, i.e. a third party, not a stakeholder
- It must be certainly one strong and rich organization or entity
- Everyone follow the same workflow, same guidelines, standard security
- An independent body should carry out development, maintenance and management of item bank.

Security

23. How can confidentiality be ensured?

- Check whether the database has been hacked
- Anticipate what happens if questions are out on the street
- Is always a problem, everything depends on trust; Database must be well secured; Do not publish questions; Confidentiality agreement
- Only members have access
- Statements should not be public and should not be shared with trainers. Security of distribution of the item bank to the sites in the different participating countries.

24. What privacy matters need consideration?

- The EU rules as laid down in the AVG (General Data Protection Regulation (GDPR) in the European context)
- anonymity of candidates
- For example, to be able to compare the performance of a task in different languages, anonymised candidate data is important, as are the answers given in the context of psychometric analyses. The privacy of candidate data is particularly important as a national problem.

Final remarks

25. Do you have anything to add?

- There is an idea in Germany that MC is always easy, just memorize a few things and the questions are always simple. Expectations must be raised that MC exams are indeed 'serious work' and therefore not inferior to an oral exam.
- It is a pity that this topic is taken up only just yet.
- Remote testing

- I am enthusiastic. There must be a good plan.
- It is a pity that this plan comes at a point in time when we have already invested a lot.

26. Do you have questions / issues …?

- What is the timeline of the research?

Annex 4: Future perspectives; a full test platform

INTRODUCTION

This Annex contains a concise description of a so called 'test platform' in a mature stage. A test platform is an accumulation of functional parts used to construct, administer and analyse tests in a professional setting. In the current phase of this project, it is not recommended to implement a full test platform. Instead, this annex could be seen as a description of a set of possible long-term goals and the connections between them. The end result would be a fully functioning modular set of tooling in service of the complete cycle of testing, which can be tailored to CESNI's needs in professionalization and standardization of tests on a European level.

A TEST PLATFORM DIAGRAM

To start with the end result, the diagram in figure 1 shows the different layers in the test platform from the perspective of functionality. The best way to explain is to do a build-up from the inside out.

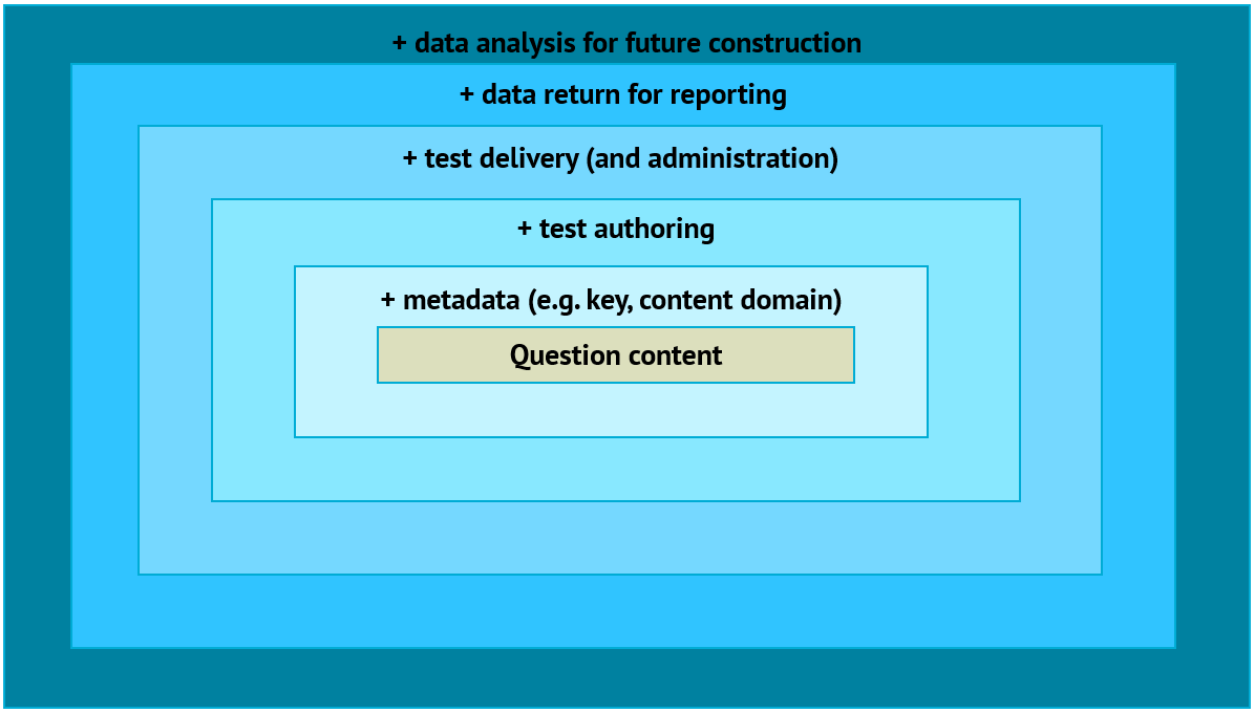


Figure 1. the different layers in a test platform

Question content

At the heart of the test platform is the question or item content. This represents the part of an item the candidate is presented with when taking a test. In a multiple choice test, this could be background information, a question and four alternatives.

Metadata

In order to categorize, distinguish, sort or in other ways efficiently handle and maintain a collection of items, a layer of metadata should be part of an item bank. Metadata about the items could for example include multiple choice key, content domain, construction period, item language, a measure of exposure, etc. In CESNI's case, the use of metadata is an important ingredient to keep an item bank in line with changes in rules and regulations.

Test authoring

Coming from the current situation of decentralized test construction, having a test platform to author standardized tests or test parts could help CESNI's goals tremendously. From the interviews conducted

with stakeholders, it became apparent that most but not all of the tested curriculum is generic. One could imagine a test administration where a generic test part would be authored on the central level, after which a decentralized test leader would add a test part specific to their region or test candidate base.

Test delivery
For the near future, this is where the test platform starts to outscope the question CESNI is currently facing. Aside from decentralized test construction, the administration process of the test is currently also conducted in a decentralized manner. Including test delivery in a test platform helps centralize part of the test administration conditions. From the perspective of the test administrator, this would mean less emphasis on organizing the test, as it is handled in part by a (digital) test delivery service. From the perspective of the test candidate, this would eliminate the situation in which the various organizations involved in the test all impact the test difficulty to different extents.

Data return for reporting
After a test administration, a transformation of information occurs: the responses of the candidate are scored, aggregated and converted into meaningful test results. Ultimately, the question to be answered in this phase is: did the candidate pass or fail? In a mature test platform, this process is automated and standardized to a large extent. Again, the cost of organizing the test chain for the test administrator is lowered. Also, standardization occurs on the level of assessing the responses: a set of responses from a test candidate in Utrecht, the Netherlands holds the exact same result value as the same set from a test candidate in Loznica, Serbia. All while improving scalability of test administrations to a large extent.

Data analysis for future construction
This final functional layer of a test platform is a key ingredient for continuously improving the entire test cycle. Test data collected over time could be used in analyses to feed the future item and test construction process, to monitor and maintain the item bank and to evaluate the connection between the test and the test goals in many ways. Periodic analyses of test data tend to drive a large step forward in test quality. For CESNI’s case, special attention should be paid to the multilingual character of the candidate base. Great care should be taken to ensure that the language of a candidate and test do not influence the probability of passing the test.

Annex 5: Example of question and metadata

An example of a question and metadata is taken from the ADN[24] catalogue of questions. It is an example of a very simple database structure. Adding metadata, or explicitly stating relations, may add to the complexity of the database.

Number	Source	Response	Remarks	Dealt with on
General				
Objective 1				
110 01.0-01	Agreement	B		19.09.2018

Examination objective 1: General

Number	*Source*	*Correct answer*
110 01.0-01	Agreement	B
	What is the abbreviation for the European Agreement concerning the International Carriage of Dangerous Goods by Inland Waterways?	
	A AITMD	
	B ADN	
	C ADR	
	D RID	

Question

The question content is usually stored in a database as follows:

Question	What is the abbreviation for the European Agreement concerning the International Carriage of Dangerous Goods by Inland Waterways?
Alternative A	AITMD
Alternative B	ADN
Alternative C	ADR
Alternative D	RID

Metadata

The metadata are usually stored as follows:

Number (Unique Identifier)	110 01.0-01
Key Answer	B
Examination objective	1 General
Source *(or knowledge to be assessed)*	Agreement
Remarks	
Dealt with on	19.09.2018

[24] The European Agreement concerning the International Carriage of Dangerous Goods by Inland Waterways

Annex 6: Example of question and metadata

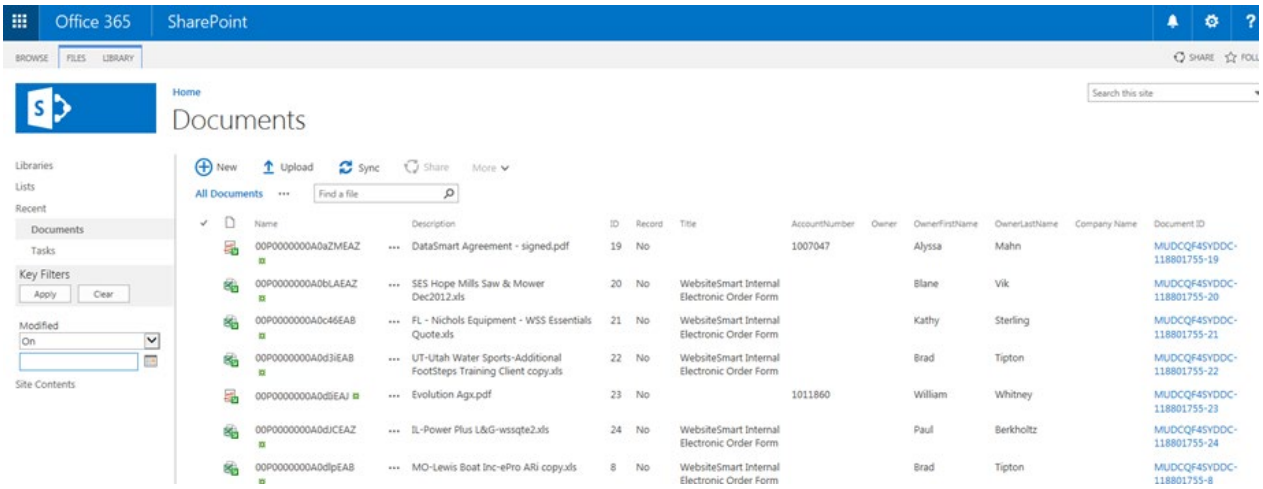
SharePoint offers the user the opportunity to store information in *lists* or in *libraries*. Lists are similar to database tables (or Excel tables). Libraries are a special type of lists, meant for archiving documents. Both lists and libraries can be customized. The availability of standard filters and customized views make it relatively easy to select and present information. Access and editing right can be assigned based on roles.

Both lists and libraries can be configured for use as a question database, but the main differences are:

List	**Document library**
All content is stored in columns of the record. (only structured data*)	Content is stored in a document, metadata of the document are shown as columns. (combination of structured and unstructured data)
Plain text	Rich text (e.g. styling, images,..)
Versioning: only major versions	Versioning: major and minor versions
	Check-in/check-out mechanism
	Approval workflow
	Document preview
	Custom document(content) types

** it is possible to attach documents to a record, but SharePoint does not offer options for document management such as in the document library.*

SharePoint hence allows for building a centralized item bank. The preparation work consists of defining the table/metadata structure and roles for access. SharePoint produces forms for entering data. The main choice to be made is the choice between using a *List* or a *Document Library.* The latter has the advantages that it supports co-editing of the item content better than the former, and that it allows to use styling/typesetting options. A (selection of) items can be exchanged with other Office applications such as Excel, or a combination of MSWord (item content) and Excel (item metadata). SharePoint does not support test assembly or other stages in the test cycle.



SharePoint document library

More information on SharePoint

SharePoint List

https://support.microsoft.com/en-us/office/introduction-to-lists-0a1c3ace-def0-44af-b225-cfa8d92c52d7

SharePoint Library

https://support.microsoft.com/en-us/office/what-is-a-document-library-3b5976dd-65cf-4c9e-bf5a-713c10ca2872

https://support.microsoft.com/en-us/office/introduction-to-libraries-7d4221d9-8fb9-40d5-8441-2374c84b5e26

Cito
P.O. Box 1034
6801 MG Arnhem
The Netherlands
T +31 26 352 13 27
international@cito.nl

Photo: Ron Steemers