



EUROPEAN COMMISSION

Horizon 2020 - Work Programme 2014-2015
LEIT – Information and Communication Technologies
Content technologies and information management
Strategic Objective ICT 15–2014: "Big Data and Open Data Innovation and take-up"

Technical background notes

This working document is intended to provide background information and technical commentary on the Horizon 2020 Work programme 2014-2015 Strategic Objective ICT 15-2014: "Big Data and Open Data Innovation and take-up".

The aim of the Strategic Objective ICT 15-2014 is to develop technologies that would increase the efficiency of all EU companies and organisations that need to manage vast amounts of data and in particular the competitiveness of EU enterprises.

For further information:

Unit Data Value Chain (G3)

European Commission – Communications Networks, Content and Technology

CNECT-G3@ec.europa.eu

For information on H2020, WP 2014-2015 LEIT-ICT:

- Horizon2020, Introduction: http://ec.europa.eu/research/participants/portal/doc/call/h2020/common/1587757-05_leit-5_intro_wp_2014-2015_en.pdf
- Work Programme 2014-2015 LEIT ICT: http://ec.europa.eu/research/participants/portal/doc/call/h2020/common/1587758-05i_ict_wp_2014-2015_en.pdf
- General Annexes: http://ec.europa.eu/research/participants/portal/doc/call/h2020/common/1587809-18_general_annexes_wp2014-2015_en.pdf

For information on previous calls on "Big Data" and "Open Data": http://cordis.europa.eu/fp7/ict/content-knowledge/home_en.html

Technical background notes for Horizon 2020, Strategic Objective ICT-15 2014 "Big Data and Open Data innovation and take-up"

DG CONNECT/G3

CNECT-G3@ec.europa.eu

http://cordis.europa.eu/fp7/ict/content-knowledge/home_en.html

This document is intended to provide background information and technical commentary on Strategic Objective ICT-15 2014 of the Horizon 2020 programme. The official text of the objective (including the timeline and procedures for applications) has been published as part of the 2014-15 Horizon 2020 work programme.

http://ec.europa.eu/research/participants/portal/doc/call/h2020/common/1587758-05i_ict_wp_2014-2015_en.pdf (pages 36-38)

The official text is the only legally binding source of information on the strategic objective. Should any inconsistency between the present explanatory document and the official text be detected it is always to be resolved in favour of the work programme text.

Motivation of the objective and scope of this document

The official text of the work programme states that this objective concentrates on:

"The activities supported under this topic address the general technological and systemic data challenges that concern entire value chains and/or bridge across borders, languages, industries and sectors. The aim is to improve the ability of European companies to build innovative multilingual data products and services, in order to turn large data volumes into semantically interoperable data assets and knowledge. The horizontal activities within LEIT on data, relevant for a wide range of sectors, will be complemented in the H2020 Societal Challenges by data-related activities addressing specific areas.." (Challenge: Content technologies and information management)

The motivation behind this objective is to develop technologies that would increase the efficiency of all EU companies and organisations that need to manage vast amounts of data and in particular the competitiveness of EU enterprises.

To quote again from the official text of the objective, if this is accomplished, the expected impact is:

- *Enhanced access to and value generation on (public and private sector) open data resulting in hundreds of multilingual applications reusing tens of billions of open data records used by millions of European citizens.*
- *Viable cross-border, cross-lingual and cross-sector data supply chains involving hundreds of European actors in a robust and growing ecosystem capable of generating sizable revenues for all the actors involved and SMEs in particular.*

- *Tens of business-ready innovative data analytics solutions deployed by European companies in global markets.*
- *Availability of deployable educational material for data scientists and data workers and thousands of European data professionals trained in state-of-the-art data analytics technologies and capable of (co)operating in cross-border, cross-lingual and cross-sector European data supply chains.*
- *Effective networking and consolidation of Big Data user and contributor communities, technology providers and other relevant stakeholders across all challenges and across the three pillars of Horizon 2020.*

This will require addressing at least all the following aspects:

- **usability:** the systems developed should be engineered so as to be usable by people whose primary occupation is not research or software development but rather running a business or an organisation. It is thus extremely important that usability be taken as a foremost concern at every step of software development and validation.
- **innovation:** the systems developed must be able to improve the processes of existing businesses or open new business opportunities. For this reason, it is extremely important that commercial companies with clear exploitation plans be at the core of consortia and define business induced technology needs to be met. Contrast this with consortia driven by researchers and technology developers adding commercial partners as use cases for technologies that are not originally motivated by a business analysis.
- **robustness:** the systems developed must be designed under the assumption that the data handled may be missing, corrupted or inconsistent; furthermore, they must be engineered to be able to be deployed outside the lab in typical operating conditions (including memory, storage and network failures).

Notice that the objective is articulated in two target outcomes (individually discussed below) so clearly distinct from one other that each proposal is expected to address exactly one of them. In general, it is important to understand that proposal evaluation will be based strictly on technical merit, management and impact potential. Please see the appendix to this document for a list of points that must be addressed to avoid submitting mediocre proposals.

a.1) Data Incubator for SMEs

Development of services based on the use of available data, particularly from public bodies, is specifically required for the first part of theme a) of the objective, which reads:

- *One collaborative project establishing a European open data integration and reuse incubator for SMEs to foster the development of open data supply chains and to educate and assist new users. Proposals are expected to:*
 - *Identify significant opportunities to establish supply chains for products and services, based on open data resources;*

- *Attract the participation of European companies willing to contribute some of their own data assets as open data for experimentation or to integrate open data with their own private data as the basis for innovative applications.*
- *Attract and manage SMEs interested in business or technology innovation in particular on open data.*
- *Link to and reuse data from the European Union Open Data Portal or other local, regional or national Open Data portals, as well as to the CEF programme.*
- *Where appropriate, link to and expand the activities of existing national/regional open data incubators.*

The action may involve financial support to third parties in line with the conditions set out in Part K of the General Annexes. The consortium will define the process of SME selection for which financial support will be granted. Minimum 70% of the EU funding requested by the proposal should be allocated to this purpose.

Here we provide additional details on the mechanics of the exercise and some guidance on some design principles that should inform proposals (and, later, the one, selected, project) responding to this part of the call.

What is the Data Incubator (DI)?

The objective of this call is to make sure that technical obstacles to the reuse of European open data and integration with private (typically enterprise) data are removed to the extent possible.

While a growing number of regional, national and international administrations (the European Commission among them) are making available growing amounts of open data, the number of applications (and the number of users of those applications) based on the reuse of those data have arguably not been growing at the same rate, despite excellent national and European initiatives in this direction.¹

It is thus possible that open data of good quality and potential interest to some end user goes unutilised due to some obstacle of administrative or technical nature. DI is designed to create an environment in which those obstacles are systematically removed and Small and Medium sized Enterprises (SME) are offered various forms of support (including financial) to pursue their data integration and reuse ideas.

In this respect, DI's high level objectives are quite similar to those of the recently launched UK's Open Data Institute² (ODI) with the difference that while ODI is focused on promoting

¹ <http://www.etalab.gouv.fr/m/article-105550957.html> <http://www.appsforitaly.org/en/>
<https://joinup.ec.europa.eu/community/ods/description>

² <http://theodi.org>

the reuse of UK open data, DI does the same for data from all over the EU and also emphasizes the integration of commercial/private data with open data assets.

How will DI help logistically?

The objective of DI is to allow SMEs to concentrate on developing and testing their data based application. Thus DI will be tasked with building and maintaining data and infrastructure for SMEs to deploy their prototypes.

An SME approved for work within DI will find at DI all the open data they need to develop their application plus a computing environment where they could integrate with other data assets and deploy their prototype application for testing.

If the data required by the SME is known to exist in principle as open data but is not yet available at DI, it will be DI's task to obtain it from the publisher and perform whatever reasonable transformations (e.g. format conversions, linking to other datasets) are required for the SME to proceed with their idea.

In order to 'prime the pump' of the open data available on its platform, DI will engage the relevant actors on both the data demand and supply directions.

An important novel aspects of the DI with respect to what was requested in a previous call is its additional focus on the integration between open data and proprietary data which is the focus of a recent (October 2013) McKinsey report³. In this scenario, an SME would approach the DI with an idea to extract value from the integration of open data with data which they own or data made available to them for experimentation/prototyping purposes by some other enterprise. For example, a visualization SME could approach the DI with an idea to integrate publicly available weather information with farmers' insurance claims data made available under appropriate non-disclosure conditions by a large insurance company.

This hypothetical scenario reveals an important requirement for the design of the DI: in order to foster this type of open/proprietary data integration, the DI must be able to provide the kind of security and privacy measures that would allow a large insurance company to feel comfortable with such integration experiments (for example in the form of tamper-resistant and fully auditable data access patterns).

How will DI help SMEs get in and out of the DI environment?

DI will be required to assist SMEs in being admitted to the DI environment and in planning their exit.

³

http://www.mckinsey.com/insights/business_technology/open_data_unlocking_innovation_and_performance_with_liquid_information

On the admission end, it will be DI's responsibility to advertise calls for proposal submissions by SMEs. This means that it will be DI's responsibility to use social networks, conferences and any other relevant means in order to spread its message where promising software development SMEs (even some who might not naturally have thought of themselves as open data re-users) are naturally found.

On the exit end, it will be DI's responsibility to educate its graduate SMEs on product development strategies and connect them with previous graduates but, most importantly, additional opportunities for funding, from venture capital to regional development funds. It is thus important that DI builds and grows a solid network of relevant contacts.

How will DI function in practice?

DI will formally exist as a large collaborative project, a project funded as a result of a competitive call for proposals submitted by a consortium satisfying the requirements for participation of Horizon 2020.

Such proposals will not be about building applications based on the reuse of open data but rather about building the DI environment in which others (and specifically SMEs) will do so. Proposals will be evaluated based on the credibility of the consortium's plans for providing all the forms of support described above.

The consortium will spend only at most 30% of the funds it receives to set up and operate the DI environment.

The consortium will disburse the majority of the funds (at least 70%) to fund mini-proposals of duration between six and twelve months submitted by SMEs in response to broadly advertised calls. Based on the amount of funding commonly offered by seed venture capital for efforts of this scope, it is expected that funding for each such mini-proposal could range between 50,000 and 150,000 Euro.

The proposed life span of the entire incubator must be appropriately motivated in the proposal in order to ensure the most effective deployment of several batches of mini-experiments. An indicative figure of 36 months is a reasonable first estimate but alternative durations are acceptable when appropriately motivated.

a.2) Collaborative projects: innovation and technology transfer

The text of the call states

Collaborative projects focused on innovation and technology transfer in multilingual data harvesting and analytics solutions and services. The projects should have a cross-sectorial, cross-border and/or cross-lingual scope, and take into account the users' and societal perspectives. The driver in consortia should be a core of companies dedicated to focused activities with a clear business perspective with verifiable milestones and market validation.

The overarching objective of this part of the objective is to foster the emergence of data value chains across Europe. A data value chain is a value chain⁴ where most of the participants have data resources as an important factor of production⁵ and produce as their output or consume as additional inputs other data resources or data technologies (e.g. database management systems, visualization toolkits, data mining or machine learning frameworks, etc...) or services (e.g. recommendation systems, navigation systems, etc...).

Projects funded under this objective are expected to directly contribute to these chains by means of the creation of data products and services that the project participants will bring to market in a sustainable fashion (as opposed to lab prototypes unfit to operate in demanding, cross border, business environments).

The even more general strategic point is that well developed value chains have been shown to be predictive of the development prospects of economic areas⁶ (through the recombination of underlying technologies and skills in pursuit of new business opportunities). For this reason, particularly valued will be solutions that enable data value chains across more than one industrial domain, EU country and language.

In the service of these high level goals, this objective is calling for data innovation and technology transfer that creates links across sectors, across national economies and across languages.

An example of innovation across sectors might be collection of data in one domain (e.g. weather) for the purpose of offering a service in a different domain (e.g. insurance for farmers).

Another example might be the development of very focused technology (for example FPGA⁷ based, real time, complex event processing⁸) in the service of two very different industry sectors such as finance and traffic management.

The examples listed above are purely indicative and do not reflect any preference for these topics as compared to many other topics that would be appropriate for this objective.

What is important, however, and constitutes a significant difference between the Horizon 2020 programme and the previous Framework Programmes is that consortia responding to this objective should be led by companies with similar or complementary business objectives (as opposed to academic teams). This has two very practical consequences:

⁴ http://en.wikipedia.org/wiki/Value_chain

⁵ http://en.wikipedia.org/wiki/Production_factor

⁶ http://en.wikipedia.org/wiki/The_Product_Space

⁷ <http://en.wikipedia.org/wiki/Fpga>

⁸ http://en.wikipedia.org/wiki/Complex_event_processing

1. the role of academic partners and software developers in the consortium is to satisfy a crisply defined business need identified by the commercial partners (as opposed to pursuing otherwise legitimate scientific curiosity or technological experimentation)
2. the commercial partners are required to give a convincing explanation of their business perspective. More than it was the case in the previous Framework Programmes, they are required to describe the competitive and technology landscape of their respective domains of activities, identify current bottlenecks or new business opportunities and describe plans for market validation

Projects should pay particular attention to issues of data quality management, i.e. the ability to harvest and extract value from data as they exist as a result of noisy, unreliable processes often beyond the control of the data user (as opposed to how data resources would look like in the ideal case of a controlled environment that guarantees accurate data modelling and complete compliance with data formats). The costs of data curation should be realistically estimated and, if necessary, also addressed by means of technological solutions that are highly usable and compatible with realistic business processes.

Once again, the realities of a multi-lingual cross-border European data-based business environment should be realistically addressed at all levels, from the encoding and management of data according to the appropriate character set to the interpretation of data assets cross languages by means of multilingual thesauri or ontologies.

b.1) Coordination and Support Actions: reuse of data assets

The text of the call states

To lay the foundation for effective exchange and reuse of data assets (including those controlled by the data subject) across: industry sectors, national boundaries and language barriers, public and private sectors. Proposals are expected to:

- *Define the legal/contractual framework that would foster the exchange of data assets and set up pilots of a self-sustaining data market;*
- *Attract and involve players from all parts of the data value chain and representing different sectors and markets;*
- *Implement a close clustering mechanism with projects arising from the last bullet point of activity a), involving them in experiments, data reuse pilots, business case workshops etc. and taking input from them in designing the legal framework and infrastructure.*

Consistently with the funding instrument indicated (Coordination and Support Actions), proposals responding to this part of the strategic objective are not expected to develop software or data assets, but rather to maximise the value of those who do so by identifying the framework conditions that would enable a functioning European data economy.

Proposals are allowed, but not required, to address all the three points discussed below. A proposal addressing in a very credible and effective way only one of the points below will be preferred to one addressing very superficially two or more.

A first desired line of activities is the study of the legal and economic conditions that would foster the creation of a functioning European data economy. A pre-requisite for legal analyses will be a thorough understanding of the EU data protection legislation⁹ and the attendant Europe-wide debate. As for economic analyses, a useful example of a research effort of the level of quality expected under this objective is the Data EcoSystem project at the University of Washington¹⁰. Needless to say, proposals that will be able to relate convincingly the findings of their legal/contractual analysis to those of their economic analysis will be preferred (quality of work being equal) to those who pursue these two lines of work independently.

Proposals are encouraged to analyse systemic factors such as the role of data assets and services in the European product space¹¹ or the effects of 'data gravity'¹² on the European cloud computing industrial landscape. These last two are only examples of the level of intellectual ambition and rigour expected of successful proposals and should by no means restrict the scope of analysis and modelling. In other words, what is expected is the development of models capable of making verifiable/falsifiable predictions that could later be checked on the basis of empirical data. What is not encouraged, on the other hand, is the submission of proposals that are vague or superficial, i.e. that do not specify the research methods to be used and that simply restate the problem of the absence of well-developed European data markets in terms of anecdotal remarks without indicating how it will be tackled.

In this context, it may be worth noting that the 2011-12 ICT work programme of Framework Programme 7, contained an objective (objective 4.1.a Bootstrapping a data economy¹³) designed

To foster reuse of digital content resources by providing guarantees and fair incentives for their creators and maintainers. This includes the creation of data exchanges or commons whose quality (breadth, timeliness, temporal qualification, ..) and value increases with the number of users and the feedback and validation they contribute. It also includes mechanisms for aggregating demand, thus stimulating the creation of additional resources and services.

⁹ http://ec.europa.eu/justice/data-protection/index_en.htm

¹⁰ <http://cloud-data-pricing.cs.washington.edu>

¹¹ http://en.wikipedia.org/wiki/The_Product_Space

¹² <http://datagravity.org>

¹³ http://cordis.europa.eu/fp7/ict/content-knowledge/fp7-chsme-excerpts_en.html

A useful exercise for those who are considering submitting proposals responding to this first point will be to acquaint themselves with the results obtained so far by the group of projects that were funded as a result of this call in order to identify lessons learned or problems identified in establishing well-functioning data markets¹⁴. Alternative/additional sources of empirical evidence on what does or does not work in establishing data markets will be equally important in order to ground proposals within the context of what is observable. By contrast, proposals that do not demonstrate a critical appreciation of past and present data market initiatives in Europe or elsewhere will be putting themselves at a considerable disadvantage.

b.2) Coordination and Support Actions: skills

The text of the call states

To contribute to capacity-building by designing and coordinating a network of European skills centres for big data analytics technologies and business development. The network is expected to identify knowledge/skills gaps in the European industrial landscape and produce effective learning curricula and documentation to train large numbers of European data analysts and business developers, capable of (co)operating across national borders on the basis of a common vision and methodology.

Although the goal of this point is rather straightforward, it is worth reviewing its underlying intent and how it will inform the evaluation of submitted proposal submitted. The overall objective is to make sure that Europe is adequately supplied with individuals who have the right set of skills to contribute to the (fast growing) data sector of the economy.

This means that proposals responding to this theme should do at least the following:

1. provide a motivated (and, preferably, quantitative) analysis of which European industrial sectors are now or are soon to become data intensive;
2. identify for each such sector the specific set of skills that are required for an organisation to be efficient and successful (for certain domains this might be data extraction and curation, for certain others statistical analysis, for others still very large scale graph analytics, visualization, data streaming, etc...);

¹⁴ These are: <http://www.eurofit-project.eu> <http://vista-tv.eu> <http://code-research.eu>
<http://www.biopoolproject.eu> <http://www.sopcawind.eu> <http://www.fusepool.eu>
<http://www.plan4business.eu> <http://www.smespire.eu> <http://www.gnss-gapfiller.eu>
<http://www.simplefleet.eu> <http://www.dopa-project.eu>

3. motivate the shortage of skill (preferably on a quantitative basis). i.e. explain why the number of skilled individuals required does or will soon exceed their current availability;
4. propose a set of learning resources to be developed, complete with a detailed process for quality assurance, coverage, modularity (i.e. how various learning resources will complement each other);
5. propose a plan to deploy those materials, preferably in environments where their effectiveness can itself be subject to rigorous, data-based, analysis (for example, in a Massive Open Online Course¹⁵)

Notice how this requires a deep understanding and, possibly, a close cooperation with European employers, possibly mediated by trade associations, industry groups, chambers of commerce, etc...

Notice also how steps 1) - 5) differ from an approach where developers of a certain data technology propose to produce and disseminate educational resources for their preferred technology, without a credible analysis as to why that specific technology would be important for the European industrial landscape nor a plan to ensure that data practitioners (and not just university students) will be able to profit from those materials. This second approach is clearly not encouraged and the evaluators of the proposals submitted will be instructed accordingly.

Part of the objective is teaching a generation of European data practitioner to cooperate across borders. This has two important consequences:

6. the educational materials will need to be appropriately multilingual. In this context, 'appropriately' means that, if the analysis in points 1) and 2) above determines that a certain data-bound industrial sector is very important in a certain EU country, educational materials for the relevant technologies will need to be produced in the language of that country (unless one can prove that most people in that country are also completely comfortable with a different language)
7. in order to foster a cross-border mindset, educational materials are encouraged to cross-reuse and build on EU data freely available from various European data portals and, in particular, <http://open-data.europa.eu> and <http://publicdata.eu> .

b.3) Coordination and Support Actions: data aggregator

The text of the call states

¹⁵ <http://en.wikipedia.org/wiki/Mooc>

To create a Big Data integrator platform with the objective to coordinate and consolidate relevant technology and user communities in any actions supported in Horizon 2020 addressing or making use of Big Data.

The operations of well-established cloud computing operators reveal a completely predictable pattern: when programming languages and technology stacks become widely used, it becomes important (and a point of competitiveness) for a cloud computing environment to support them and offer them to their tenants, much in the way heating, electricity and connectivity are offered as part of a standard rental contract for office space.

This theme invites proposals to survey the Big Data needs of all components of the Horizon 2020 funding programme (but, in particular, its Social Challenges¹⁶), identify components that would be needed by several objectives and assemble them into cloud-enabled software stacks.

Given the amount of funding available for this effort, it is not expected that the funded project would actually run an end-to-end computing centre on behalf of the communities it is designed to serve. Rather, to pursue the analogy introduced above, it is designed to work like a model office space run by a real estate agency: a model office space is not meant to actually host tenants nor to provide full time electricity, heating, etc... but it would be capable of so doing if this became necessary and so it can convince perspective tenants that if they decided to rent all the relevant infrastructure would be in place.

It is expected that proposals responding to this theme would survey the two orthogonal dimensions of

1. domain specific data assets and technology and
2. generic, enabling technologies

As a concrete example of what this means, one could observe the current operations of the Open Phacts platform¹⁷. In this environment, domain specific data assets are pharmaceutical databases and domain specific technologies are pharmacology search engines and visualization environments; generic, enabling technologies include the linked data infrastructure (RDF store, SPARQL endpoint) on which the platform is based.

It is further to be expected that certain basic technologies and data assets would be of interest for more than one community. For example, geospatial data assets and mapping technologies are likely to be of interest for all Horizon 2020 activities that are concerned with entities or activities that exist in physical space. Similarly, graph databases and visualization technologies and systems for minting and reusing unique identifiers are likely to be of interest in any domain (from industrial logistics to food chains) where it is important to keep track of

¹⁶ http://ec.europa.eu/research/horizon2020/index_en.cfm?pg=better-society

¹⁷ <http://www.openphacts.org>

very large numbers of individuals and their relationships. Scalable workflow managers or database management systems such as Hadoop, or Cassandra are going to be needed by domains where the very scale of data requires sharding.

It is important to notice that the community building aspect of the work solicited is just as important as the technical one: it will be the responsibility of the funded consortium to maintain meaningful working relationships with researchers and developers working in other areas of the Horizon 2020 work programme to ensure that their needs are fully understood and that the platform assembled can meet those needs (and grow to accommodate emerging ones) in a way that would more or less ensure the (re)use of the platform in future developments of the Horizon 2020 funding programme.

Appendix: a list of questions that proposals must answer in order to fulfil strategic objective ICT-15 2014

This appendix contains a list of simple questions that a consortium should ask about the proposal to be submitted. If the proposal as submitted does not contain a clear answer to the majority of the relevant questions for the various outcomes it places itself at a serious disadvantage in a very competitive selection process (because the evaluators of the proposals will be specifically instructed to look for the answers to these and other questions).

a.1) Integrated Open Data Incubator for SMEs

1. How is the consortium putting the needs of SMEs (as opposed to those of the core partners) at the centre of its planning?
2. How is the consortium planning to maximise the amount of resources available to SMEs and minimise the amount used by the core partners of the consortium?
3. How is the consortium planning to attract promising SMEs, assist them in the process of submitting a proposal to the planned calls and assist the select SMEs in complying with the requirements of participation in a H2020 project?
4. Does the consortium have a strategy concerning the data resources that it plans to curate for the benefit of SMEs? Are there data domains that mutually increase each other's value if well curated?
5. Does the consortium have clear plans (and credible cost projections) for the computational infrastructure that it plans to make available to SMEs for the development of application prototypes?
6. Does the consortium have clear plans to help SMEs migrate from the prototype environment to independent applications?
7. Does the consortium have a clustering strategy to encourage the participation of SMEs with mutually value enhancing application ideas?
8. What is the consortium's specific strategy to connect SMEs with additional sources of support (e.g. venture capital)? Does the consortium have quantitative targets in this respect? If so, how will it monitor its progress?

a.2) Innovation and technology transfer

1. What is the business or process improvement opportunity that is being pursued? What data is being supplied in support of this opportunity (e.g. number of customers that would buy the technology if it can be made to meet certain price/performance parameter) and how could one, after the end of the project, determine if the opportunity has been met or not?
2. What exactly is the innovative aspect of the work proposed? Does it consist in using existing technology in an unexpected environment or in its extension for a

specific use case (as a dedicated geospatial module might be for a database management system)?

3. If the project is about technology transfer, what exactly is being transferred, and why can't the market be expected to apply these technology as a natural development? In other words, what makes the technology transfer risky and what is the project proposing to mitigate that risk?
4. If the proposal describes cross domain activities, what is the industrial case for those activities? What are the links of the relevant value chain and which of these links will benefit (and to what extent) from the activities proposed?
5. If the proposal is cross-country or cross language, what motivates this choice? What makes this combination of countries/languages a particularly promising domain to innovate in data technologies?

b.1) Coordinating and Support Actions: reuse of data assets

1. What factual evidence (in the form of datasets, reports) will be used as the basis for legal, economic and systemic analysis on the establishment of data markets in Europe?
2. If the proposal intends to gather for its analysis factual evidence that is not today available, what data will it gather, at what cost and according to what methodology?
3. If the proposal intends to gather for its analysis factual evidence that is not today available, will it later make it available as open data for others to study and reuse if they see fit? If so, what data formats will it use to ensure maximal reusability of the data so produced?
4. How does the proposal intend to investigate constrains that are based on the structure of the markets involved (for example, the fact that certain data activities that are technologically and legally possible do not take place simply because they would run against the commercial interests of existing players or their efforts to defend a competitive position)?
5. How will the proposal test its recommendation against the history of data reuse activities (or lack of activity) in Europe or elsewhere? In other words, how will it avoid recommending things that have been shown not work elsewhere (or show that conditions have now changed)?

b.2) Coordinating and Support Actions: skills

1. What factual evidence will the proposal use or collect to justify the need to develop educational materials for a given data asset or technology?
2. How will the proposal determine that such resources do not exist elsewhere or are very unlikely to be developed elsewhere?
3. How will the proposal collect the honest views of the industrial sectors it purports to serve? How will those views be validated after the end of the project?

4. How will the proposal track the number of learners who use their learning resources?
5. How will the proposal assess the quality of the learning resources it has produced in terms of the skill gains of the end users (as opposed to internal quality review processes)?
6. In which European languages will the learning materials be produced and why?

b.3) Coordinating and Support Actions: data aggregator

1. What process will the proposal use to gather and validate the requirements of the communities it intends to serve? How will it maximise the chances that the platform will in fact be later used by the intended communities?
2. How will the proposal ensure that the platform will be modular and encourage reuse of cross-cutting basic technologies (as opposed to duplicating efforts by building separated vertical stacks for each community addressed)?