Enabling procurement data value chains for economic development, demand management, competitive markets and vendor intelligence

Deliverable 4.1

Existing Visualisation and Interaction Components

<table>
<thead>
<tr>
<th>Delivery Date:</th>
<th>June 28, 2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s):</td>
<td>Tom Blount (SOTON), Elena Simperl (SOTON), Laura Koesten (SOTON)</td>
</tr>
<tr>
<td>Dissemination Level:</td>
<td>Public</td>
</tr>
<tr>
<td>Work package:</td>
<td>WP4: Interaction design and storytelling</td>
</tr>
<tr>
<td>Version:</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Project co-funded under Horizon 2020 Research and Innovation EU programme under Grant agreement no. 780247
Document Metadata

Quality assurers and contributors

<table>
<thead>
<tr>
<th>Quality assurer(s):</th>
<th>Ian Makgill (OO)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contributor(s):</td>
<td></td>
</tr>
</tbody>
</table>

Version History

<table>
<thead>
<tr>
<th>Date</th>
<th>Version</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>04/06/2018</td>
<td>0.1</td>
<td>Initial draft for internal review</td>
</tr>
<tr>
<td>12/06/2018</td>
<td>0.2</td>
<td>Expanded ideation summary, adding additional information on Data &amp; Processes, Users, and Communication of Data</td>
</tr>
<tr>
<td>19/06/2018</td>
<td>0.3</td>
<td>Added additional study for Competitive Markets &amp; Procurement Intelligence</td>
</tr>
<tr>
<td>25/06/2018</td>
<td>0.4</td>
<td>Described how the success criteria of the user studies in §3 relates to the conceptual framework in D4.2</td>
</tr>
<tr>
<td>28/06/2018</td>
<td>1.0</td>
<td>Submission-ready version</td>
</tr>
</tbody>
</table>
Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Figures</td>
<td>4</td>
</tr>
<tr>
<td>Definitions</td>
<td>5</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>6</td>
</tr>
<tr>
<td>2 Ideation and Survey</td>
<td>7</td>
</tr>
<tr>
<td>2.1 Methodology</td>
<td>7</td>
</tr>
<tr>
<td>2.2 Outcome Summary</td>
<td>8</td>
</tr>
<tr>
<td>2.2.1 TBFY</td>
<td>8</td>
</tr>
<tr>
<td>2.2.2 Data and Processes</td>
<td>9</td>
</tr>
<tr>
<td>2.2.3 Users</td>
<td>9</td>
</tr>
<tr>
<td>2.2.4 Communication of Data</td>
<td>10</td>
</tr>
<tr>
<td>3 User Studies Plan</td>
<td>11</td>
</tr>
<tr>
<td>3.1 Economic Development</td>
<td>11</td>
</tr>
<tr>
<td>3.1.1 Search Presentation Study</td>
<td>11</td>
</tr>
<tr>
<td>3.1.1.1 Research Questions</td>
<td>11</td>
</tr>
<tr>
<td>3.1.1.2 Methodology</td>
<td>11</td>
</tr>
<tr>
<td>3.1.1.3 Proposed Timeline</td>
<td>12</td>
</tr>
<tr>
<td>3.1.2 Search Facet Study</td>
<td>12</td>
</tr>
<tr>
<td>3.1.2.1 Research Questions</td>
<td>12</td>
</tr>
<tr>
<td>3.1.2.2 Methodology</td>
<td>13</td>
</tr>
<tr>
<td>3.1.2.3 Proposed Timeline</td>
<td>13</td>
</tr>
<tr>
<td>3.2 Demand Management</td>
<td>13</td>
</tr>
<tr>
<td>3.2.1 Trend Recognition and Prediction Study</td>
<td>13</td>
</tr>
<tr>
<td>3.2.1.1 Research Questions</td>
<td>13</td>
</tr>
<tr>
<td>3.2.1.2 Methodology</td>
<td>14</td>
</tr>
<tr>
<td>3.2.1.3 Proposed Timeline</td>
<td>14</td>
</tr>
<tr>
<td>3.3 Competitive Markets &amp; Procurement Intelligence</td>
<td>15</td>
</tr>
<tr>
<td>3.3.1 Search Experience Study</td>
<td>15</td>
</tr>
<tr>
<td>3.3.1.1 Research Questions</td>
<td>15</td>
</tr>
<tr>
<td>3.3.1.2 Methodology</td>
<td>15</td>
</tr>
<tr>
<td>3.3.1.3 Proposed Timeline</td>
<td>15</td>
</tr>
<tr>
<td>3.3.2 Collusion Detection Study</td>
<td>16</td>
</tr>
</tbody>
</table>
D4.1– Existing Visualisation and Interaction Components
Dissemination Level: Public

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.3.2.1</td>
<td>Research Questions</td>
<td>16</td>
</tr>
<tr>
<td>3.3.2.2</td>
<td>Methodology</td>
<td>16</td>
</tr>
<tr>
<td>3.3.2.3</td>
<td>Proposed Timeline</td>
<td>17</td>
</tr>
<tr>
<td>3.4</td>
<td>Initial Gantt Chart</td>
<td>17</td>
</tr>
<tr>
<td>4</td>
<td>Conclusions</td>
<td>19</td>
</tr>
<tr>
<td>5</td>
<td>References</td>
<td>20</td>
</tr>
<tr>
<td>A</td>
<td>Information Sheet for Interview Participants</td>
<td>21</td>
</tr>
<tr>
<td>A.1</td>
<td>Human-Data Interaction</td>
<td>21</td>
</tr>
<tr>
<td>A.1.1</td>
<td>Example</td>
<td>21</td>
</tr>
<tr>
<td>A.2</td>
<td>Requirements analysis</td>
<td>23</td>
</tr>
</tbody>
</table>
List of Figures

3.1  OO’s Spend Network tool ................................................................. 12
3.2  OO’s OpenOpps Dashboard ............................................................. 14
3.3  ZGZ’s Electronic Town Hall ............................................................ 16
3.4  Initial Gantt chart for the proposed studies ........................................ 18

A.1  “Steroids or Not, the Pursuit Is On”, New York Times analysis of baseball player Barry Bonds’ performance ................................................................. 22
A.2  “Norse Attack Map”, a live visualisation of cyber-attack data (http://map.norsecorp.com) 22
A.3  “Parable of the Polygons”, a set of interactive visualisations using simulation to teach about segregation and diversity (http://ncase.me/polygons/) ......................... 23
Definitions

Abbreviations

**OCDS**: Open Contract Data Standard

**PPROC**: Public Procurement Ontology
1 Introduction

This deliverable describes part of the initial ideation process of T4.1, specifically the survey of consortium members, making use of the conceptual framework for describing dimensions of data visualisation based on the review of background literature and media described in D4.2. The aim of this was to determine the current state of existing processes, user needs and challenges, visualisations and overall goals within the consortium. These elements will be described in more detail in the upcoming deliverable D6.1. This deliverable also proposes a plan for user studies to be carried out in the coming months.

This document is of interest for the TBFY consortium to be informed about the ongoing process of developing suitable visualisation and interaction components to meet the specific user needs arising from TBFY, and the future visualisation and interaction prototypes (D4.4). Outside of TBFY, the deliverable is useful for future projects and initiatives intending to make use of human-data interaction and data visualisation.

The first part of this deliverable describes the methodology of the ideation process, and a summary of the key outcomes. The second part of the deliverable provides a plan for future user studies over the course of the project, to examine the outcomes in further detail, and to aid the evaluation of the visualisation and interaction components developed during T4.2. The results of these studies will be presented alongside D4.4.
2 Ideation and Survey

The goal of T4.1 is ideation of visualisations and interactions for presenting data; the main aim of this task is to understand the planned business cases and their outcomes, and the data storytelling elements that are relevant for the business cases. Here we present a survey and examples of the existing data visualisation and interactions methods used by the partners. This will be followed up and extended in future deliverables (D4.4) to include mock-ups and software prototypes.

2.1 Methodology

In conjunction with the conceptual framework described in D4.2, we conducted one-on-one interviews with the business and technical partners to determine the existing needs of stakeholders and their users, how these are currently fulfilled, and the stakeholders’ future goals. These were semi-structured interviews which allowed for freedom for discussion, and aimed to have a duration of no more than 60 minutes. We asked the following questions, which were provided to participants prior to the interview (along with supplementary background material on human-data interaction, reproduced in Appendix A):

**TBFY Project**

- What are your goals as an organisation:
  - in general terms?
  - in terms of project outcomes?

- How does (or how could) your data impact the four TBFY scenarios (search, sensemaking, visual analytics, decision making etc.)?

- What are your plans for WP6? (for business-case partners)

**Data and Processes**

- What is the pipeline of your process (i.e. how do accumulate and aggregate raw published data to a structured form)?
  - Where do your source your data?
  - What types of data do you collect/prioritise?
  - What technologies do you currently use to store/represent your knowledge graph?
• What manual curation is currently required:
  – to acquire data?
  – to maintain it?

Users
• What are the goals and data-needs of your users?
  – (What do users use your system for? Why?)
• What challenges do you face in supporting your users?
• What actions do you expect your users to take after viewing your data?
• Do you have channels of contact for existing users (for future user studies etc.)?

Communication of Data
• What existing visualisations/dashboards do you use?
  – Please provide screenshots if possible
• How do you communicate complex information to your users? Such as:
  – uncertainty in the data?
  – quality of the data?
  – provenance of the data?
  – other metadata?
• Do you use live data, or an aggregation?

The following partners participated in the interview process: CERVED, OO, MJU and JSI. Additional informal discussions were had with OC, UPM, and ZGZ. The interview results were processed using thematic analysis and an open coding system, to discover similar attributes discussed by the participants. A summary of the results for each of the interview sections follows.

2.2 Outcome Summary

2.2.1 TBFY
The interviewed participants highlighted two key long-term goals that guided their interest within the project. Firstly, collecting and cataloguing financial and tender data, to support buyers and suppliers (CERVED, OO). Secondly, to increase the transparency of the procurement process, to build public trust, and prevent the formation of bidding cartels (MJU, JSI). The different users involved in each of these scenarios are described in further detail below.
2.2.2 Data and Processes

Consortium members currently source data from multiple third-party providers, both purchased from commercial vendors, and acquired from open-data services (one particularly frequently cited source was the “Tenders Electronic Daily” portal\(^1\)), and official government gazettes. The data itself is retrieved through API access, bulk-downloads, and scraping of web-based portals. Due to the number of disparate data providers (OO alone draws data from between 300-400 sources, both national and international), there is an inherent variety in the formats of the obtainable data, including XML, RDF and JSON, with no unified standard. However there is a concerted effort within the project to move towards the Open Contract Data Standard\(^2\) (OCDS) and the Public Procurement Ontology\(^3\) (PPROC).

The primary cost in terms of processing and manual curation is in the initial set-up of tools to access the data, reason over it and link it with other data (CERVED described a semi-automated, doubt-resolution process, used for linking entities) and to transform it to a uniform format. However, there is also the (sometimes substantial) cost of on-going maintenance and management, cause by data-providers changing hosting location, publication format, or otherwise altering the process of access, as well as maintaining and validating mappings between, for example, redundant data sources. Partly due to this, OO specifically made the case for prioritising data sources based on those with the highest cost-benefit ratio; that is, those that are most easily accessible, and provide the highest throughput of data, with the lowest risk of regular change in the access provided.

In some cases (such as those described by MJU) it is the suppliers and/or buyers themselves that are responsible for providing records of this data (and are indeed mandated to do so). However, this can lead to issues surrounding the quality of data they provide (outlined in further detail below), particularly when the providers are unsure of what data is required from them, or are under-resourced to provide it.

2.2.3 Users

There were a number of key users identified by the participants. Naturally, procurement managers were cited as a central type of user in most systems. Their role is publish invitations to tender, using data to make informed and effective purchasing decisions at the most competitive rates, and ensuring oversight in terms of understanding payments being made, and in terms of which departments have contracts with which suppliers.

Equally naturally, suppliers responding to tenders are another key group of users. They have a number of goals when using data, including understanding the current state of the market based on both region and domain, and planning a long-term bidding strategy through being aware of upcoming tenders (and the expiration of existing contracts).

In addition, audit managers (alongside associated fraud prevention bodies) were cited as an important user group (particularly by MJU) in their role of detecting fraud, collusion, and bid rigging among suppliers through the awareness of, for example, links between the management structure of suppliers and the tendering department. The National Audit Office\(^4\) (NAO) in particular references the lack of available contract data to aid in this regard, in a recent report on the liquidation of management and construction contractor Carillion (National Audit Office, 2018).

Finally, the general public was identified as another key group of users by participants for which openness and transparency (again, particularly MJU) were a critical factor of their business case. JSI highlighted that

\(^1\)http://ted.europa.eu/TED/main/HomePage.do
\(^2\)http://standard.open-contracting.org/latest/en/
\(^3\)http://contsem.unizar.es/def/sector-publico/pproc.html
\(^4\)https://www.nao.org.uk
users from the general public are often more interested in data that proves directly relevant to them: for example, prioritising regional over national or international data. OO also highlighted the role of journalists and researchers in analysing and disseminating this information to the public.

The common challenge outlined by all participants was the issue of data quality and completeness, both in terms of how to manage representing and reasoning over flawed data. The quality of gathered data can be low for a number of reasons. Primarily, there are datasets that are missing (either in whole or in part), exist but remain unpublished, or are otherwise unavailable, which leads to gaps in the aggregation and visualisations. In addition, information on licensing agreements associated with (re-)publishing data is often lacking, as is consistent and dependable meta data. Finally, the data itself can be incorrect or misleading, which can occur either by accident (in which the data provider makes an innocent mistake), or through the direct action of the data provider (either because of a miscommunication of what data needed to be represented, or the malicious altering of the published data to hide the evidence of collusion or fraud). This leads to problems around reasoning over this data (such as linking duplicate entities with minor differences) and representation (such as how to inform users that the data is subject to uncertainty).

2.2.4 Communication of Data

CERVED described the primary means of providing their data for the project as being through APIs, and therefore unreliant on visualisation technologies; however they described instead other products provided by their organisation which did feature visualisations. The majority of participants described their data being as close to live as possible, usually updating on a daily basis.

Representing complex data types such as multi-faceted contract data (containing a variety of data dimensions such as start-date, end-date, value, category/domain, etc.), presents a challenge: while it is simple to create different types of visualisations of each of these dimensions, it is not currently clear which, if any, would be of most use to end-users and would likely depend heavily on their task and expected outcome. An approach towards studying this in further detail is discussed further in §3.

In terms of addressing the challenges described in §2.2.3 (particularly the issue of data quality and completeness), OO highlighted that while it is important to ensure that users are aware of the uncertainty inherent in the published data, it can be difficult to highlight this naturally in the visualisations, and that the simplest way can simply be to include a blanket header with caveats. Another way of addressing this issue is visualising the provenance of the data. However, for particular business cases there is also a risk with directly linking to the provenance of their data sources: when a business model relies on aggregating data for the convenience of their users, this could be undermined by providing users direct access to the original sources. A potential means of addressing this challenge is to provide users with tools to report data that is known to be inaccurate or, depending on the circumstances (such as if the user group can be considered trustworthy), allow them to make or suggest changes themselves, similar to the “Wiki” model.
3 User Studies Plan

To progress the ideation process, we propose a number of user studies to evaluate the existing visualisation and interaction (including search) technology used by consortium partners, by examining the effects of data presentation on task effectiveness. As the ideation and development process continues, the design of these studies may be updated (and new studies may be proposed) as needed, to fully explore the different scenarios. The results of these studies will be presented in a future deliverable (D4.4).

3.1 Economic Development

3.1.1 Search Presentation Study

This study aims to explore how the presentation of search results affects a user’s judgement of a dataset for a particular task (and their confidence of the same), with specific relation to procurement data.

3.1.1.1 Research Questions

RQ1: What is the impact (in terms of usability, memorability, understanding, and aesthetics) of the presentation of search results (e.g. title/metadata only, textual summary, or dataset excerpt) when searching for procurement data?

3.1.1.2 Methodology

Participants must be data-literate (that is, they must be familiar with searching for, and using, data to complete tasks); ideally, participants should be data professionals used to working with tender data, either as a procurer or supplier. Participants will be asked a series of questions to gauge both their demographic background and their data-literacy.

Participants will then be presented with a data-related task that they will be instructed to complete, such as finding the current supplier of a particular contract, or finding a list of all contracts in a given domain, below a given value. However, the exact nature of the task must be informed by prior research in the area (such as the work by Wehrend and Lewis (1990)), or in-depth interviews with data professionals familiar with working with tender data, and justified accordingly.

Participants will then presented with a search interface (as an example, the current interface used by OO’s SpendNetwork tool is shown in Figure 3.1) and asked to complete the task by applying queries and browsing the results. Information about participants’ task performance, such as search queries attempted, refinements made, and time taken to complete the task (or abort it) will be recorded. Finally, participants will be asked to report their answer for the task, and their confidence that it is correct.
To determine the suitability of each visualisation, participants’ performance will be measured in terms of the evaluation framework described in D4.2: usability, understanding, and (where appropriate) memorability. Participants will also be asked to complete a short series of questions judge their perception of the aesthetics of the visualisations (such as ease of use, clarity, and overall preference).

3.1.1.3 Proposed Timeline

M7: Initial feasibility workshop(s) with OO/OC
M7: Data generation and sourcing
M8: Submission to ethics board, task design, pilot study
M9: Rollout of full study
M10: Results analysis

3.1.2 Search Facet Study

This study aims to explore how use of advanced search features (such as facets, filters, and other refinements (Hearst, 2006; Ben-Yitzhak et al., 2008)) impact the speed and accuracy with which users complete a particular task (and their confidence of the same), with specific relation to procurement data.

3.1.2.1 Research Questions

RQ1: What is the impact (in terms of usability, memorability, understanding, and aesthetics) of advanced search features when searching for procurement data?
RQ2: What advanced search features do participants use the most when searching for procurement data?

RQ3: How do advanced search features affect participants experience of searching for procurement data?

3.1.2.2 Methodology

As before, participants must be data-literate and ideally should be data professionals familiar with working with tender data, either as a procurer or supplier. Participants will be asked a series of questions to gauge both their demographic background and their data-literacy.

Participants will then be presented with a task that they are attempting complete, such as finding the current supplier of a particular contract, or a list of all contracts in a given domain below a given value. Again, the exact nature of the task must be informed by prior research in the area, or interviews with data professionals.

Participants will be divided into experimental groups and presented with a search interface (as an example, the current interface used by OO’s OpenOpps Dashboard is shown in Figure 3.2). One group will have no access to faceted search features. Other groups will have access to one or more facets. Participants will be asked to complete the task by applying queries and browsing the results. Information about participants’ task performance, such as search features used, queries attempted, refinements made, and time taken to complete the task (or abort it) will be recorded. Finally, participants will be asked to report their answer for the task, and their confidence that it is correct.

To determine the suitability of each visualisation, participants’ performance will be measured in terms of the evaluation framework described in D4.2: usability, understanding, and (where appropriate) memorability. Participants will also be asked to complete a short series of questions judge their perception of the aesthetics of the visualisations (such as ease of use, clarity, and overall preference).

3.1.2.3 Proposed Timeline

M7: Initial feasibility workshop(s) with OO/OC

M7: Data generation and sourcing

M8: Submission to ethics board, task design, pilot study

M9: Rollout of full study

M10: Results analysis

3.2 Demand Management

3.2.1 Trend Recognition and Prediction Study

This study aims to aid the Demand Management scenario by studying the ways in which users recognise trends and apply this information to making predictions.

3.2.1.1 Research Questions

What is the impact (in terms of usability, memorability, understanding, and aesthetics) of different visualisation modes when searching for trends in procurement data?
3.2.1.2 Methodology

As before, participants must be data-literate and ideally should be data professionals familiar with working with tender data, either as a procurer or supplier. Participants will be asked a series of questions to gauge both their demographic background and their data-literacy.

Similar to the studies described above, participants will be given a data-related task to complete, this time focusing on making predictions using available data to analyse, for example, regional demand. As before, the exact nature of the task must be informed by prior research in the area, or interviews with data professionals.

Participants will be divided into experimental groups; each group will be presented with a different visualisation component. Participants will be instructed to explore the available data and complete the task. Information about participants’ task performance, such as interaction with visualisations, and time taken to complete the task (or abort it) will be recorded. Finally, participants will be asked to report their answer for the task, and their confidence that it is correct.

To determine the suitability of each visualisation, participants’ performance will be measured in terms of the evaluation framework described in D4.2: usability, understanding, and (where appropriate) memorability. Participants will also be asked to complete a short series of questions judge their perception of the aesthetics of the visualisations (such as ease of use, clarity, and overall preference).

3.2.1.3 Proposed Timeline

M7: Initial feasibility workshop(s)

M9: Data generation and sourcing
3.3 Competitive Markets & Procurement Intelligence

3.3.1 Search Experience Study

This study aims to explore what features of data search and visualisation are most effective at improving user experience with specific relation to searching for procurement data in the context of market competitiveness.

3.3.1.1 Research Questions

RQ1: *What is the impact (in terms of usability, memorability, understanding, and aesthetics) of search interface when searching for procurement data relating to competitive markets?*

3.3.1.2 Methodology

Participants must be data-literate and ideally should be data professionals familiar with working with tender data, either as a procurer or supplier. Participants will be asked a series of questions to gauge both their demographic background and their data-literacy.

As before, participants will be given a data-related task to complete, this time focusing on assessing areas for potential competition, such as finding imminently expiring contracts in a certain sector. However, once again, the exact nature of the task must be informed by prior research in the area, or interviews with data professionals.

Participants will then presented with a search interface and asked to complete the task by applying queries and browsing the results (as an example, the current interface used by ZGZ’s Electronic Town Hall is shown in 3.3). Information about participants' task performance, such as search queries attempted, refinements made, and time taken to complete the task (or abort it) will be recorded. Finally, participants will be asked to report their answer for the task, and their confidence that it is correct.

To determine the suitability of each visualisation, participants' performance will be measured in terms of the evaluation framework described in D4.2: usability, understanding, and (where appropriate) memorability. Participants will also be asked to complete a short series of questions judge their perception of the aesthetics of the visualisations (such as ease of use, clarity, and overall preference).

3.3.1.3 Proposed Timeline

M7: Initial feasibility workshop(s) with ZGZ
M10: Data generation and sourcing
M11: Submission to ethics board, task design, pilot study
M12: Rollout of full study
M13: Results analysis
3.3.2 Collusion Detection Study

This study relates to the detection of collusion, bid-rigging and fraud by suppliers, and would be the first step towards determining whether this type of activity can be gamified, to engage the general public with the work carried out. The key element here is to firstly determine whether it is possible for non-experts to recognise markers of collusion when presented with a visualisation of the relevant data with sufficient accuracy to at least highlight a case for further investigation, before then exploring how different visualisation and interaction techniques can be applied to support this activity.

3.3.2.1 Research Questions

RQ1: Can non-experts detect potential collusion between suppliers when presented with a visualisation of tender data?

RQ2: How does the accuracy of non-experts at detecting collusion compare with that of experts?

RQ3a: How does the type of visualisations used affect the accuracy of predictions?

RQ3b: Which type of visualisation gives the best result?

3.3.2.2 Methodology

Initially, participants will be shown a set of instructions providing basic information about the task, alongside positive and negative examples of evidence related to collusion. Participants will be asked demographic background questions, before being shown a series of visualisations of (either anonymised or fabricated) data, and asked to make a judgement as to whether they believe it is evidence of collusion, or not.
The data shown to participants (and the visualisations used) must be informed by relevant literature (e.g. (Porter and Zona, 1993), (Conley and Decarolis, 2016)), as well as consortium members that regularly manage data for the purposes of auditing and fraud-detection (e.g. CERVED and MJU), to ensure that it has appropriate markers of collusion and/or fraud, and is also what could be considered a realistic dataset.

The design of this task (specifically the lack of required skills and the volume of data to be processed) lends itself to the use of microtask-crowdsourcing to reach a large number of non-expert participants at relatively low cost.

Further studies will need to provide a more exploratory interface, such that participants must search for the evidence required, to test a more realistic detection scenario. These studies can be repeated with expert-participants, and the change in accuracy due to different visualisations and tools compared.

### 3.3.2.3 Proposed Timeline

- M7: Initial feasibility workshop(s) with CERVED/MJU
- M10: Data generation and sourcing
- M11: Submission to ethics board, task design, pilot study
- M13: Rollout of full study
- M13: Results analysis

### 3.4 Initial Gantt Chart
D4.1– Existing Visualisation and Interaction Components

Dissemination Level: Public

Figure 3.4: Initial Gantt chart for the proposed studies
4 Conclusions

Informed by the conceptual frameworks described in D4.2, this deliverable described the initial ideation process of T4.1, specifically the survey of consortium members and a proposal for future user studies to guide the design, development, and evaluation process of the interaction and visualisation components to be developed during T4.2 (and to be presented in D4.4 and D4.6).

There are a number of factors that influence the types of visualisations required to support the consortium: notably, there are key audiences with differing needs that must be supported. There are also common challenges to face when representing complex procurement data, particularly with regards to how to represent the quality of data being displayed.

These factors will be taken into account during T4.2 and, in conjunction with the conceptual framework presented in D4.2, will be used to support the development and evaluation of future visualisation and components.
5 References


Appendix A  Information Sheet for Interview Participants

A.1 Human-Data Interaction

The topic of Human-Data Interaction (HDI) seeks to analyse the decisions and actions users take when in the role of data-analysts, and as data-subjects (Haddadi et al., 2013; Mortier et al., 2013). Simply being presented with raw data, or unfiltered algorithm results, is overwhelming to most users; to aid them in making sense of it, it is often necessary to find a way to present the data in a way that is appropriate for the audience (such as using the data to tell a story (Knaflic, 2015)), to keep the audience interested and engaged (Ma et al., 2012). Visualisations can be used to support a narrative and reduce the cognitive effort required to understand complex information. Increasingly, authors requiring the means to communicate data to a wide audience (such as online journalists) are integrating visualisations into their narratives, or even telling the story solely through visualisation (Segel and Heer, 2010).

A.1.1 Example

Examples include curated narrative visualisations (such as the New York Times’ visualisation of baseball player Barry Bonds’ performance, shown in Fig. A.1), real-time visualisation of “live” streaming data (such as the Norse Attack Map, shown in Fig. A.2), interactive and/or “playful” visualisations and simulations (such as the “Parable of the Polygons”, shown in Fig. A.3) and reports that are automatically generated from raw data using natural language generation (NLG) (such as those used by Narrative Science).
D4.1– Existing Visualisation and Interaction Components

Dissemination Level: Public

Figure A.1: “Steroids or Not, the Pursuit Is On”, New York Times analysis of baseball player Barry Bonds’ performance

Figure A.2: “Norse Attack Map”, a live visualisation of cyber-attack data (http://map.norsecorp.com)
Figure A.3: “Parable of the Polygons”, a set of interactive visualisations using simulation to teach about segregation and diversity (http://ncase.me/polygons/)

A.2 Requirements analysis

The goal of T4.1 is ideation of visualisations and interactions for presenting data; the main aim of this is to understand the planned business cases and their outcomes, and the data storytelling elements that are relevant for the business cases.

To achieve this we will be conducting a requirements analysis of the stakeholders and business cases; we intend to conduct one-on-one interviews with the business and technical partners to determine the existing needs of stakeholders and their users, how these are currently fulfilled, and the stakeholders’ future goals.

This will be a semi-structured interview, with the stakeholder of the relevant business and technical partners. The interview will have freedom for discussion, and should take no more than 60 minutes. We will be asking the following questions:
TBFY Project

• What are your goals as an organisation:
  – in general terms?
  – in terms of project outcomes?

• How does (or how could) your data impact the four TBFY scenarios (search, sensemaking, visual analytics, decision making etc.)?

• What are your plans for WP6? (for business-case partners)

Data and Processes

• What is the pipeline of your process (i.e. how do accumulate and aggregate raw published data to a structured form)?
  – Where do your source your data?
  – What types of data do you collect/prioritise?
  – What technologies do you currently use to store/represent your knowledge graph?

• What manual curation is currently required:
  – to acquire data?
  – to maintain it?

Users

• What are the goals and data-needs of your users?
  – (What do users use your system for? Why?)

• What challenges do you face in supporting your users?

• What actions do you expect your users to take, after viewing your data?

• Do you have channels of contact for existing users (for future user-studies etc.)?

Communication of Data

• What existing visualisations/dashboards do you use?
  – Please provide screenshots if possible

• How do you communicate complex information to your users? Such as:
  – uncertainty in the data?
  – quality of the data?
  – provenance of the data?
  – other metadata?

• Do you use live data, or an aggregation?