

# DELIVERABLE 2.3

## ParCos Curator



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 872500.



**“ParCos – Participatory Communication of Science”  
A HORIZON 2020 RESEARCH AND INNOVATION ACTION**

**Consortium:** Lappeenranta-Lahden teknillinen yliopisto (FI, coordinator), Katholieke Universiteit Leuven (BE), Vlaamse Radio- en Televisieomroeporganisatie (BE), and Knowledge West Media Centre LBG (UK).

**Webpage:** <https://parcos-project.eu>

**Duration:** 1/2020 – 12/2022

**Grant:** H2020-872500 (Call H2020-SwafS-2019-1)

**Contact (co-ordinator):**

Asst. Professor Antti Knutas & Asst. Professor Annika Wolff

LUT University

e-mail: [parcos.project@lut.fi](mailto:parcos.project@lut.fi)

**Disclaimer:** This document’s contents are not intended to replace consultation of any applicable legal sources or the necessary advice of a legal expert, where appropriate. All information in this document is provided "as is" and no guarantee or warranty is given that the information is fit for any particular purpose. The user, therefore, uses the information at its sole risk and liability. For the avoidance of all doubts, the European Commission has no liability in respect of this document, which is merely representing the authors' view.

**DESCRIPTION OF THE DELIVERABLE**

<b>Overview</b>	<b>Details</b>
Authors	Md Tanvir Hasan, LUT (Finland) Natasha Tylosky, LUT (Finland) Annika Wolff, LUT (Finland)
Reviewers	Antti Knutas, LUT (Finland)
Number of Deliverable	2.3
Title of Deliverable	ParCos Curator
License	CC BY 4.0, see <a href="https://creativecommons.org/licenses/by/4.0/">https://creativecommons.org/licenses/by/4.0/</a>
Attribution	CC BY 4.0 ParCos, <a href="http://parcos-project.eu">http://parcos-project.eu</a> H2020-872500
Dissemination Level	Public
Contractual delivery date	2021-04-30
To be cited as	Hasan T., Tylosky N., and Wolff A. (2021). Deliverable name, deliverable 2.3 of the Horizon 2020 project ParCos, EC grant agreement no 872500, Lappeenranta, Finland.

## SUMMARY

---

This deliverable describes the ParCos data curator and how the principles for data curation were derived. This includes analysing museum practice and drawing parallels towards data curation. Also, proposing the use of narrative principles to guide data curation and new metadata. Finally, reviewing some of the key research activities on playful, participatory and interactive data curation.

## TABLE OF CONTENTS

---

1	Introduction – the role of curation in ParCos.....	6
1.1	Requirements for the ParCos curator .....	7
2	Curation as museum practice .....	8
2.1	Identification, procurement and management .....	8
2.2	Selection, organization and presentation for an exhibition (with a focus on museum stories) .....	9
3	Curation as data practice .....	9
3.1	Metadata aspects.....	11
3.2	Visual aspects .....	14
3.2.1	Data Comics .....	14
3.2.2	Brooke Leave Home .....	15
3.2.3	Datacatcher.....	15
3.2.4	PlayBat .....	15
3.3	Combining narrative metadata with data visualisation.....	16
4	Data literacy.....	17
5	Listing the Principles .....	17
6	Turning principles into practical ideas.....	18
7	Case studies and curation .....	18
7.1	Rethink remake recycle.....	18
7.2	Veden Äärellä .....	19
8	Summary.....	20
9	References .....	20
10	Appendix A – curator cards.....	22

# 1 INTRODUCTION – THE ROLE OF CURATION IN PARCOS

**Participatory science** is used to refer to the different types of involvement people may have in developing and undertaking scientific research, such as within citizen science activities. In ParCos we use the existing Bristol framework (D2.1) to scaffold the process of identifying problems of shared concern, and framing and conducting different types of experimentation to solve these problems. Unlike the previous work on the Bristol Framework, we are most interested within the ParCos project in how to make the **communication** of science outcomes to also be a participatory process. *Thus, a key objective of the ParCos project is the creation of **participatory science stories**, which focus on the types of involvement that people may have with scientific processes once scientific outcomes are communicated.*

Science communication has in the past typically been quite static. Scientific outcomes are conveyed to an audience, through different means such as news media, tweets, blogs, documentaries and similar. The evidence and data on which the communication is based may be made available separately, for example through an open research data portal, but this link is rarely made clear. Instead, some **curated snapshots of data** may be created and published along with the science story to back up the claims - but the audience have little opportunity to investigate further. Figure 1 shows a screenshot from a news site with a typical approach to telling a science story. In this case, the story is about COVID infections and vaccinations from the New York Times. In the bottom right of the first screen, there are interpretation of trends in data. This story offers the possibility to click and see data related to keywords, however like most curated data of its kind it does not do a good job of linking the data to the story. Specifically, the story is comparative: it talks about growth rate of one state compared to all others, but clicking for more details only shows details of that one state, so there is no way for the reader to check the comparison for themselves. Also, the capacity to interact with the data is limited to what is visible on the screen, or what can be reached through the hyperlinks provided.

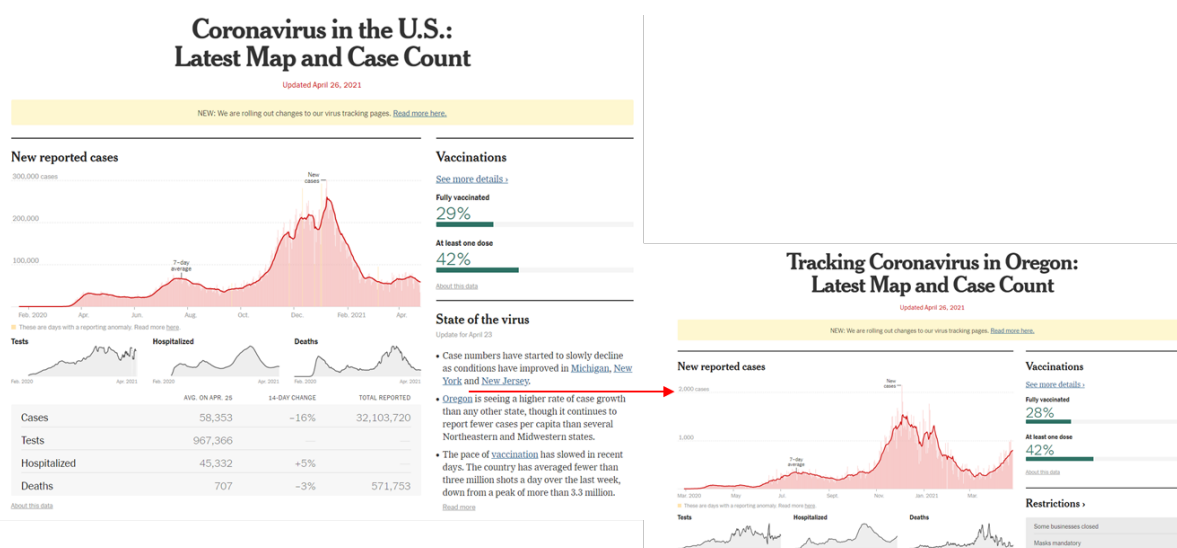


Figure 1. Typical Covid story with curated data (taken from the New York Times, 26-4-2021)

While some news sites do better in providing pointers to where original data can be found, or even in providing more interactive visualisations, there are no standard approaches to follow for curating data within a science communication.

ParCos aims to create a set of data curation principles that better explain how to curate data for science communication. In ParCos, the aim is to use curated data as *the first entry-point through which an audience may begin to engage with scientific evidence for themselves*, either validating the story that is being told, finding their own interpretation, or using it as a starting point to a new line of inquiry (figure 2). It is therefore important that curated ParCos data can lead easily to ParCos data exploration. Thus the ParCos curator is linked to both the *ParCos Data Explorer* and the *ParCos storyteller*.

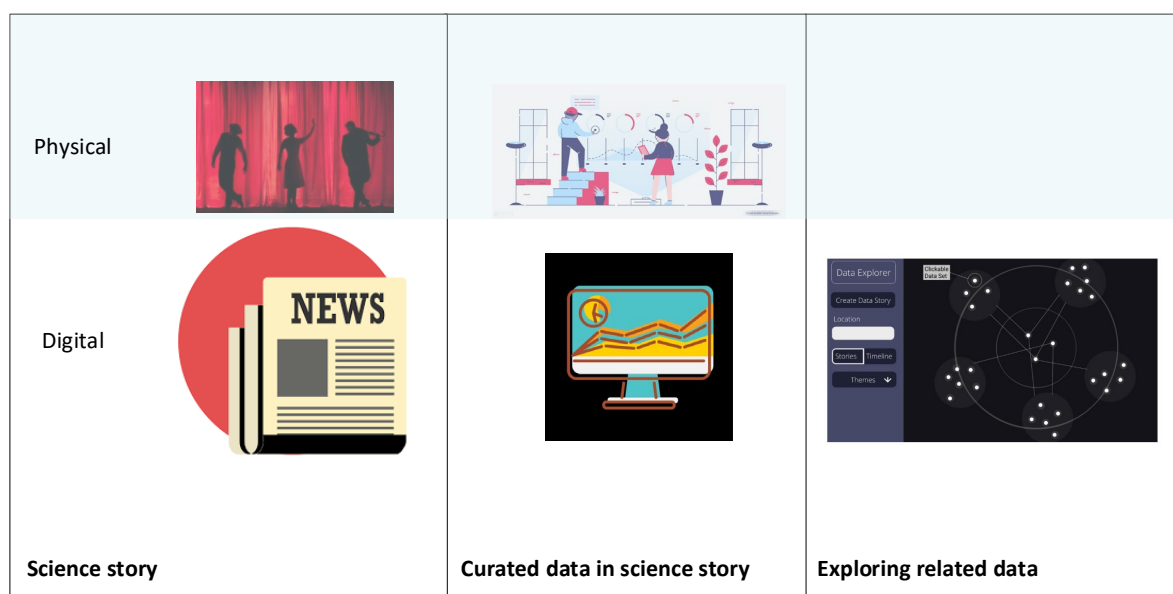


Figure 2. Showing data curation as a bridge between science story and data exploration in physical and digital worlds.

The aim of parCos is to find new ways to tell participatory science story, using popular media forms. Currently, in ParCos, this encompasses approaches such as museum exhibits, zines, visual narratives (a form of interactive visual book), live action games and a form of interactive theatre.

### 1.1 REQUIREMENTS FOR THE PARCOS CURATOR

Given the role of the ParCos curator as a link between a science story and wider data sets, it is possible to identify a number of key requirements for the ParCos curated data.

- The curated data must be **meaningful** in the context of the story being told
- It should be **easy and fun** to use, otherwise the audience will not want or be able to explore further
- Since ParCos stories are intended to be told in new ways and possibly through different media forms, it is necessary to consider that data curation may also **take place in these different media** - which in turn may be digital (e.g. interactive book), or non-digital (e.g. theatre performance or live game) - in order to fit with how the science story is being told.

- Curated data should **act as a link to the Data Explorer**, which itself supports exploration of data from beyond the scope of the original science communication. This link may be *direct* if the story is digital or *indirect* if the story is being told in a non-digital form. In this case, there is a further requirement to signify how to find and explore additional data or new perspectives on the same data.

In order to develop the data curator according whilst also taking these requirements into account, the following approaches have been taken:

1. Understanding curation as a museum practice and as a practice applied to data
2. Literature research into participatory and engaging approaches to data curation, to fit the types of scenarios outlined above
3. Rapid prototyping of data curation principles based on prior work in this area
4. Evaluation of current ParCos case study activities to identify existing practices of curation

These activities will be described in the remainder of the deliverable D2.3 as well as the outlining of the planned next steps for continued development, testing and refinement of ParCos curator. The Curator will ultimately be delivered as a set of design principles for curating data, based on the findings. These principles will be delivered in an easy to use way, such as design cards, and made available from the ParCos platform.

## 2 CURATION AS MUSEUM PRACTICE

---

The term *curation* originated as part of museum practice. The various curation roles in a museum relate to two types of practice. These are 1) the identification, procurement and management activities related to the museum collection 2) the selection, organization and presentation of museum objects for display as part of an exhibition. Activities for each include:

### 2.1 IDENTIFICATION, PROCUREMENT AND MANAGEMENT

- a) **Collect.** Identifying and collecting objects that are of interest to the museum and that are in scope of the types of things they are interested in. For example, if they are a museum of 20th century folk art they are less likely to want to collect and display 16th century renaissance paintings.
- b) **Archive.** Collecting and storing information related to the object. For example, this could include recording all known details of the object and its history, as well as current information on where the object has been loaned/borrowed/displayed and also how it links to other artefacts either in the museum or elsewhere.
- c) **Preserve.** To ensure the preservation of items in the museum collection, including understanding how items should be stored and implications of different ways of interacting with artefacts. Examples include strategies to prevent damage that may be caused by handling or using artefacts, or placing them in certain atmospheric conditions, such as damp or light. Also, understanding any dangers posed by the object towards the public, such as capacity to cause physical injury, to offend or be culturally insensitive.



## 2.2 SELECTION, ORGANIZATION AND PRESENTATION FOR AN EXHIBITION (WITH A FOCUS ON MUSEUM STORIES)

- a) **Sensemaking and story construction:** to organize museum objects for the public with respect to an overarching story. In this role, the act of curating means selecting and presenting objects for an audience in a way that reveals interesting stories about i) individual objects ii) across a set of objects. To view objects together to see what extra insight this brings.
- b) **Present:** To present objects in a physical or virtual space in a way that reflects the museum story. This may include determining the positioning of objects, the angles from which it is possible to view (for example a statue placed against a wall where the visitor cannot walk behind). It also includes defining the ways that people can interact with it, or alternatively how they are discouraged from interacting with it such as use of cordons, 'do not touch' signs, or placing out of reach or behind glass. It also includes understanding how objects may provoke reactions particularly if presented in certain manner or juxtaposed against other objects.
- c) **Storytell:** To convey the museum story to the exhibition visitors, including providing access to key information about individual objects as well as how they relate to the main exhibition story and to other objects. This could utilise, for example, information panels, audio tours, tour guides, AR applications, virtual gallery tours.
- d) **Visitor experience:** visitors experience the museum story in different ways. How they engage with and understand the museum story may be influenced by a number of factors: the order in which they visit objects, the objects they pay most attention to, the extent to which they read provided information or make their own interpretation, whether they use a real or virtual tour guide that may pick out alternative highlights and stories to the intended curated exhibition and finally their own knowledge, cultural background and experience (Schorch, 2013).

Curation has more recently been used to refer to the act of collecting and presenting things on social media. One example is Pinterest, where people identify things they want to 'collect' and then present them under a common theme. Such curation practices are often more lightweight than traditional curation practices of museums, for example there may be less thought given to relationships between objects or stories that can be told about them. Often the content collected is a replica or snapshot of some existing content and so a specific aspect of the original is preserved but not the item itself.

The common acts of curation in both these cases are that the objects somehow belong together under some theme, that someone (the curator) interprets them in that context - even though other groupings/contexts could be possible - and that they also show how these various things are related to each other.

## 3 CURATION AS DATA PRACTICE

---

It is possible to draw clear parallels between the curation activities in museums and data curation. Similar to museum curation of objects, **data curation** relates to the activities of managing and improving data after it's generation. The aim of data curation is to ensure its availability and reuse in future. Based on the data curation lifecycle model proposed by Digital Curation Center (DCC), "Data curation is a process of selectively implementing ongoing and systematic maintenance and management of reliable scientific data of reuse value from its generation in line with scientific data lifecycle to ensure reuse and value addition of the data,

including a series of activities such as scientific data planning, data creation or collection appraisal & selection, organization & disposal, description, transformation, storage and reuse” (Zhang and Zhao, 2017). Table 1 shows clearly this relationship between curation of museum objects and the curation of data within the context of creating participatory science stories.

*Table 1. Relating Museum curation to data curation in the context of supporting and creating participatory science stories*

	Curation Activity	Data
<b>Identification, procurement and management</b>	Collect	Identify and collect relevant data that is relevant to the theme of an (open) data repository. For example <i>open research data</i> , <i>open government data</i> and so on. This separation allows each type of repository to offer specialized support.
	Archive	Research and capture metadata, including the data provenance and also usage history ( <i>information on where it has been loaned/borrowed/displayed</i> ).
	Preserve	Understand appropriate and inappropriate contexts of use for the data, for example bounding the geographical area in which it has relevance. Consider privacy and security concerns of data, GDPR, data bias.
<b>Selection, organisation and presentation</b>	Sensemaking and story construction	Identify stories within and across datasets ( <i>using the ParCos Data Explorer, upcoming deliverable D6.2</i> ) and understand how the data evidences a particular view on a science story. Understand and visualize relationship between datasets, e.g. placing on same graph to see what extra insight this brings
	Present	Decide how to present data in the context of the participatory science story. Consider the potential emotive aspects of the data. Will the data be presented in digital or non-digital form. To what extent will it be interactive?
	Storytell	Convey the participatory science story to the audience, including deciding a) what information is needed to be shown about a data set in the context of a story b) how to make clear relationship between dataset and the overarching story c) how to show relationship between different data sets
	Visitor experience	The audience participates in the story by bringing their own interests and experiences to understanding the story and associated datasets. They build an emotional connection

		towards the story and the issues revealed via the data that evidences it.
--	--	---

There are many aspects of data curation that are important to ensure re-use. For ParCos, we mainly focus on how to improve data curation that is part of *archiving processes* and also *presentation processes* as these have the most relevance in the context of using data to support participatory science stories. Within both archiving and presentation of data, we consider the following:

1. **Metadata aspects:** good metadata descriptions support data reuse by making it easier to construct queries that will find data, as well as making it easier to judge the relevance of returned data.
2. **Visual aspects:** To facilitate data reuse it is often necessary to make a visual representation of the data, to turn something very abstract into images and graphics. Visual aspects are important for sensemaking and for presenting data to others.

### 3.1 METADATA ASPECTS

The metadata aspects of data curation are important to create data that is FAIR<sup>1</sup>. FAIR data is data that is:

- a) Findable – for both humans and computers
- b) Accessible – even where authentication and authorization processes are needed
- c) Interoperable – with the potential to integrate easily with other data
- d) Reusable – so that they can be replicated and/or combined in different settings

In ParCos, we are less interested in data discovery ‘from scratch’. Nevertheless, participatory science stories in ParCos are intended as potential entry points not just for exploring the data directly associated with the story, but for potentially finding new data from which to make sense of the story or tell new stories.

To support this, we propose that *narrative principles* can be applied to metadata-related data curation activities to make data more reusable and also findable in the context of a particular science story.

People are inherently narrative thinkers and learn to make sense of the world through stories from an early age. Narratives have certain common elements which include the *setting* (when and where the story takes place), *characters* (human or non-human) and *themes* (Chatman, 2021). These are the elements of a narrative that provide *coherence*, so that the events that are happening in any given story may move around in time or place, follow different characters, or even explore different themes, but as long as enough things stay constant it is possible to follow what is happening and how everything in the story is related.

**Data sets can be described in terms of these same narrative principles.** Data collection takes place within specific *settings* (Loukissas, 2019) and can be identified according to different *themes* as well as having been collected from or about different *characters*. All of this provides important contextual information that may help discoverability of data, may drive *data recommendation* and crucially, gives those who have not had the benefit of understanding first-hand the context of the data collection more information from which to recreate it

<sup>1</sup> <https://www.go-fair.org/fair-principles/>

somehow. These ideas are represented through the conceptual framework in Figure 3. This represents a process through which narrative principles may be applied towards describing data, organising it into loose narrative collections of related data and finally turning it into data stories through a process of applying increasing contextual interpretation.

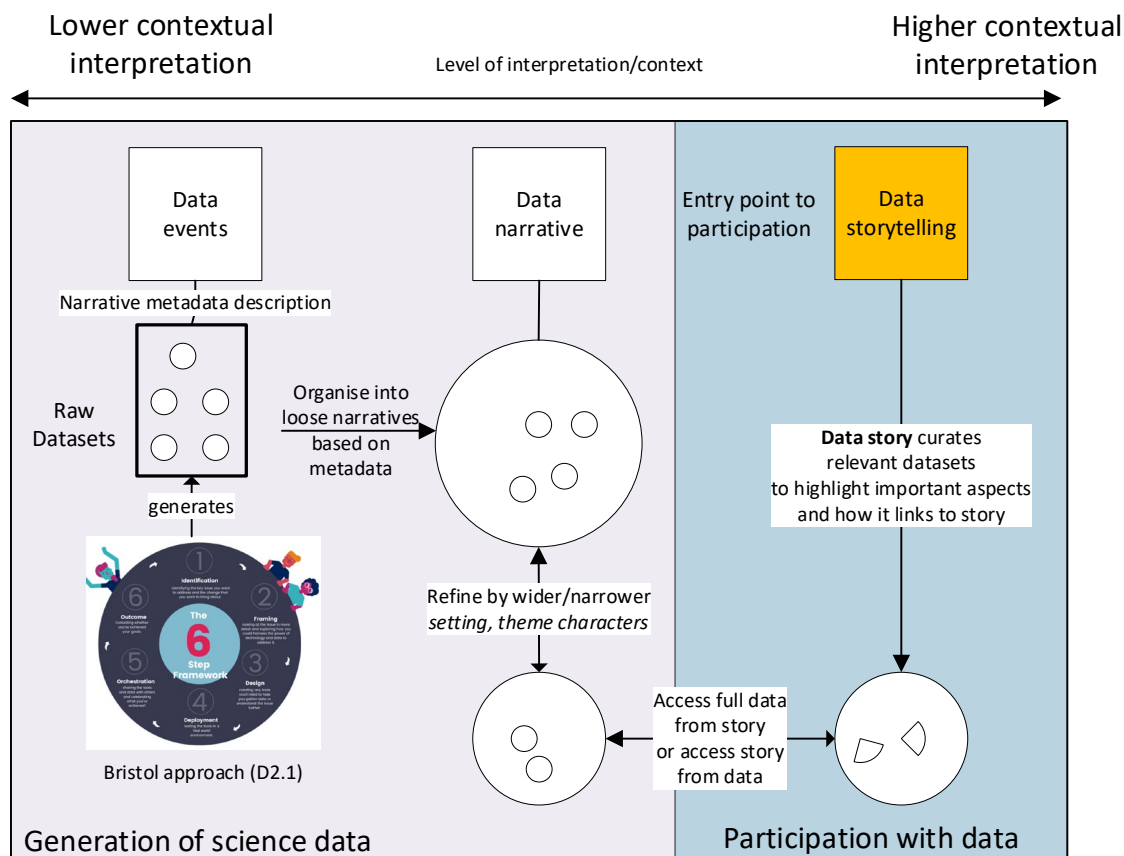


Figure 3. conceptual framework describing transformation of raw data into data narratives and data stories, with increasing levels of interpretation

The steps are as follows:

1. **Data events** A *data event* captures important information about a data sets, relating either to its *collection* or its *reuse*. **Collection events** describe the circumstance of raw data collection in terms of the setting, theme and characters represented (in addition to other standard metadata of open data). This reflects the interpretation that has already been placed on the data by choices made about the purpose and manner of its collection. **Re-use events** should capture time and place of reuse but may also include user-generated reviews and annotations (Koesten et al., 2017). To further support data re-use, *data teasers* may be created that show this metadata to help judge quality, usability and relevance at a glance. A teaser in this case is a carefully selected snapshot or view of the data that supplies the most informative information about it. **Creating data events is an archiving activity.**
2. **Data narratives:** A number of data events may be related and organised into a loose narrative, based on overlap of narrative properties. These represent data sets that *may* be useful if combined, although the choice of combination depends on the exact problem being solved. Filters may widen or narrow the scope of a narrative across

these different dimensions (e.g. expanding or narrowing the time period or themes). **Data narratives are part of sensemaking and story construction.**

3. **Data story:** A data story is defined by a user and it is *a specific interpretation of data that bounds a proposed narrative by determining the setting, theme and characters* that are relevant to that particular viewpoint across the dataset(s). A data story is therefore linked to the datasets and *specifically the parts of the data that are used in its creation*. A data story can act as an *entry point* for exploring data. By applying narrative principles in reverse, the user may start to expand the story in directions they find interesting or to discover new datasets that offer new interpretations across the data, using the ParCos Data Explorer. Thus, each time data is explored and used in a story there are more elaborations and ideas on how different data may fit together and this in itself can support new data discovery and data search. For any given dataset it may therefore be possible to see the one or more stories associated to it. **Data stories are related to both presentation and storytelling.**

Therefore, using the above framework it can be demonstrated that curating data with narrative principles may be useful at different stages of a data lifecycle, from its original collection to its use within a participatory science story.

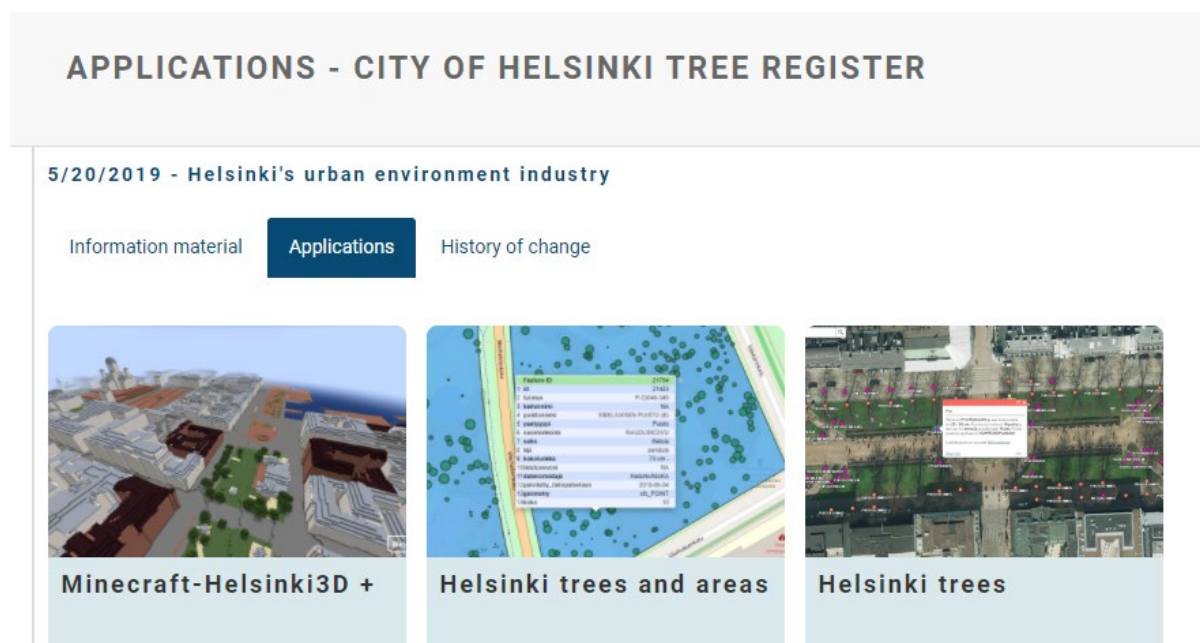


Figure 4. Screenshot from HRI.fi open data repository, showing applications that are associated with a published data set about trees (screen has been translated from Finnish using Google translate)

Figure 4 shows a visual example in use of a data curation practice within the Helsinki Region Infoshare open data portal<sup>2</sup>. Here, open data sets occasionally contain links to examples of applications in which they have been used, which is part of archival processes of capturing and showing to users the history of the data. ParCos proposes to supplement this practice by also showing if the data has been used as part of science stories.

<sup>2</sup> <https://hri.fi/data/fi/dataset/helsingin-kaupungin-puurekisteri>

To summarise the above as principles for data curation:

1. Describe data according to narrative principles of setting (time and place), themes and characters (as well as other FAIR metadata)
2. Use story (or sub-story) *settings, themes and characters* to guide what data, and what parts of data are curated for a story.
3. Create links between data and stories it is used in

## 3.2 VISUAL ASPECTS

The example shown in figure 3 demonstrates how metadata that is stored about the history of the use of a data set can be presented to a user in a visual way. In that example, a screenshot has been chosen that gives some indication of how the data has been visualized within the associated applications. Theory, technique and methods that are used to transform data into graphics are called data visualisation. Generating and visualizing data is not a new practice. Back in 18<sup>th</sup> century bar and line charts were used as form of data visualization. Still they are in regular use. There are many desktop tools available for this, such as Excel, Tableau, MATLAB, programming language R. A good data visualisation should help to improve the processing and interpretation of data for better support and understanding (Yan, Wang and Xia, 2017). In ParCos we are also concerned with visual aspects of data curation and how they support sensemaking and foster interest in deeper data discovery. Here we draw on and briefly summarise recent key literature on playful and interactive data curation, to understand how narrative principles may integrate with them and what we may learn with regards to visual aspects of data curation.

### 3.2.1 Data Comics

Data visualisation doesn't only refer to traditional methods, such as representing data as charts and graphs. Several unique techniques are being used to represent data to the public. One of these methods is the **data comic**, which use the visual language and storytelling concept from traditional comics to help explain key insights about data and make the visualization process simple (Wang, Bach and Dingwall, 2019). Bach *et al.*, 2018 describe in their paper design patterns of data comics which describe panels with a specific narrative purpose, for example to support identifying trends or making comparisons of data sets.

Another data visualization technique inspired by data comics presented in (Wang *et al.*, 2020) is **cheat sheets** which aims to present data to a wide audience. Traditional cheat sheets are piece of papers with written notes. In the past, they have been used to teach programming language. In data visualization they have been use to help designers to select chart types. The cheat sheets presented by Wang et al. are mainly inspired from infographics and comics and are aimed at helping to understand complex visualizations. A related technique presented in (Bach *et al.*, 2016) is a **graph comic** in which visual expressiveness and familiarity of comics are used to explain changes in networks to an audience. All the above research has demonstrated that the comic approach for curating and presenting data can help people without any training to understand it.

### 3.2.2 Brooke Leave Home

*Brooke Leave Home* is a film designed to support public engagement with open data that evidences the narrative, which is about what happens when someone in the UK leaves a care situation. The experience is intended to use both the data used for the main story and the data that people look at from their own location to build empathy towards the issues faced by people leaving care situations, where they may struggle with the amount of money they get to start out with. In the documentary, the story follows a girl called Brooke leaving care and it shows what happens at typical steps of this process, for example showing how much money they have to spend just buying basic supplies and how much this takes up of their available money. At critical points, viewers have access to a map where they can explore what the situation would be like in their own local area, by seeing this data in comparison to that shown in the film. In *Brooke Leave Home*, Concannon et al. (2020) have demonstrated how it is possible to both curate data and link it to a story, whilst leaving the possibility for people to participate by personalizing their view of the story through engaging with equivalent data to that which is shown in the film, but that is more relevant to their own local area. This data is accessed through an interactive map. The data is curated using simple and easy to follow visualization techniques, it is embedded into the story at key points and the way of interacting with the data is made clear to those following the video.

### 3.2.3 Datacatcher

Next, the Datacatcher was a device that scraped and curated location-based data and presented it through a device that people carried with them (Gaver et al., 2016). The data shown on the device therefore changed as people moved to different locations. The topics of the data were related to sociopolitical issues in the UK. Thus, the device might give information about average earnings, air pollution, percentage of empty properties and health of the local population. To make the data more meaningful it was presented as very short data stories, that would relate the data from the current location to one that was very close by. Gaver et al. found that while many people did not see the utility of the data it *did* capture the imagination of some, prompting them to do further data exploration. However, overall it did not work very well as a tool for fostering engagement towards the important issues the data related to. This could be because the DataCatcher left much of the interpretation of the data and the storytelling up to the person holding the device.

### 3.2.4 PlayBat

By contrast, Kaninsky et al. (2018) designed a situated interface called *PlayBat* that engaged the public with IoT 'bat call' data, with the intention to foster curiosity towards bats, which are creatures that are typically feared by many people. The IoT data was *curated* to make it more accessible to the public using different colours to summarise data which were shown on a map. The data visualization was interactive, such that a slider could be used to change the time span of the data being visualized. A narrative was used to structure the activities that people could do with the data, however, the data itself was not *interpreted* for the public and the public could not explore data beyond the boundaries set by the curation choices (they could only go 10 days back, for example).

### 3.3 COMBINING NARRATIVE METADATA WITH DATA VISUALISATION

All of these examples above - Data comics, Brooke Leave Home, DataCatcher and PlayBat – represent ways that a dataset might be curated and in some cases also how it might be embedded within a story (especially in the case of Brooke Leave Home). All of these approaches lend themselves to being framed according to the narrative principles of *setting*, *theme* and *characters*.

**Data comics** consist of a series of panels that reveal different aspects of data that is being presented, as well as providing the possibility to tell a story around them. The data comic approach is supported by templates that can be used for different types of data, for example time series data can be shown in a temporal series, *zoom* can show increasing levels of detail on a dataset, which could be used especially for geo-spatial data and contrast shows two visualisations side by side. Adding narrative principles alongside the data comic approach might be used to select what should be shown in a snapshot and this in turn can be derived from the story (or parts of the story) being told.

**Brooke leave home** does not employ any novel data visualization techniques, as it uses mainly maps, graphs and statistics - but it shows very clearly how data can be curated and embedded to evidence part of a storyline. The data in Brooke leave home appeared on the screen alongside the narrative at key moments. Whilst it was not formally created from such principles, the curated data can be related to the story according to the narrative principles of *setting*, *theme* and *characters* as a more formal way of determining at what junctures data should be embedded and what data to show.

**Datacatcher** used *setting* to determine what data was most relevant to the person holding the device. Both Brooke leave home and Datacatcher used this notion of *immediate* relevance to try to engage people further with the data, under the assumption that people are more likely to engage with and be interested about *local issues*. Datacatcher explored a number of different sub themes under a common theme of sociopolitical issues. Therefore, within a fixed *setting* (time and place) the person holding the device could explore those different themes.

**PlayBat** also situated data as close to the original *setting* as possible by placing the public display into the location from which the original data was obtained. Playbat fixed the spatial aspects of the setting but allowed people engaging with the public data display to explore backwards in time by a fixed number of days. The character (the bat) and the theme stayed constant. PlayBat was successful in evoking curiosity and perhaps even empathy towards bats.

The following data curation practices can be identified from this literature search:

1. Simplify data visualization to make it easier to read (Data comics, Data catcher, PlayBat)
2. Try to make data curations as *relevant* to the audience as possible, personalizing the data if possible (Brooke Leave Home, DataCatcher)
3. Use storytelling to frame the experiences with the data and guide the points of interaction (Brooke Leave Home, PlayBat)
4. Foster empathy through data curation (Brooke Leave Home, DataCatcher, PlayBat)



## 4 DATA LITERACY

---

Data literacy refers to the capability that people have for making sense of and using data. Data literacy principles can also be applied to data curation to make the curated data easier to use. The following data literacy principles have been identified in prior research (Wolff et al., 2019):

- 1) **Expansion:** start from a representative snapshot of a small part of data from where the audience can expand out, rather than starting with the whole data set and focusing in
- 2) **Context:** present data in a context that is relatable to the audience, for example to the local environment
- 3) **Inquiry:** Give worked examples of data inquiries and prompt ways to expand this.
- 4) **Personal Data Collection:** the best way to understand data is to go through the process of collecting it for yourself
- 5) **Foundational Competencies:** provide more support for how to ask good questions from data, rather than specific skills for data analysis which vary depending on the context
- 6) **STEAM:** combine creative with practical activities

The *context principle* is closely aligned to the goal of making data curations as relevant as possible to a user through personalization. Therefore these principles could be merged with each other. The *expansion principle* can be achieved by applying narrative principles to choose the representative snapshot that is the proposed starting point for data exploration.

## 5 LISTING THE PRINCIPLES

---

This section lists all the data curation and data literacy principles derived through the activities outlined in this deliverable.

1. The lifecycle of data curation for participatory science stories is analogous to the curation of museum objects for telling museum stories
2. Describe data according to narrative principles of setting (time and place), themes and characters
3. Use story (or sub-story) *settings, themes and characters* to guide what data, and what parts of data are curated for a story.
4. Create links between data and stories it is used in
5. Simplify data visualization to make it easier to read
6. Try to make data curations as *relatable* to the audience as possible, for example using data from the local environment, or personalizing the data if possible
7. Use storytelling to frame the experiences with the data and guide the points of interaction
8. Foster *empathy* through data curation
9. Start from a representative snapshot of a small part of data from where the audience can expand out, rather than starting with the whole data set and focusing in
10. Give *worked examples* of data inquiries and prompt ways to expand upon this towards own interests
11. Engage people with a data set that they didn't personally collect by encouraging them to collect their own data and add to it

12. Provide support for how to ask good questions from data, rather than specific skills for data analysis which vary depending on the context
13. Combine creative with practical activities

## 6 TURNING PRINCIPLES INTO PRACTICAL IDEAS

---

A small number of 'curator cards' have been created which are based on some of the principles described above and which propose different activities with data that could be embedded into participatory science stories. These are in Appendix A.

The practical activities on these cards include:

1. **Comic strip:** Make a comic strip about the dataset
2. **Historian:** build a chronological timeline out of a dataset
3. **Investigation:** start with one core 'clue' piece of data and curate data around it
4. **Mini experiment:** conduct a mini experiment by collecting and analysing data
5. **News article:** make a news article out of datasets
6. **Roleplay:** have a Q&A with an 'expert' and invite audience questions

These cards will be further refined and updated during parCos as they are utilized in case studies and made available from the ParCos platform.

## 7 CASE STUDIES AND CURATION

---

Two ParCos case studies have so far collected or selected data and used it as part of telling a science story. The following section summarises some of the approaches taken for data curation in ParCos and reflects upon how they align to some of the principles being identified in this document. In the further case study activities the ParCos curator principles will be used directly to guide data curation within participatory stories, which will give the opportunity to evaluate their use in practice.

### 7.1 RETHINK REMAKE RECYCLE

ReThink Remake Recycle explored ways to reduce household waste. This activity is being presented formally as part of D5.2 The activity began with online workshops during which data was collected about household waste and the participants came up with ideas about how to improve the situation. As one outcome of this activity, a Zine was created from data collected during online workshops around home waste recycling practices. Essentially, the data used as evidence for the zine was collected by participants of the online workshops – following the Bristol Approach - and then presented through the zine to a wider audience (this being effectively the science communication stage). In this case, the data presented in a participatory way even though the zine itself was printed material. This was done using quiz questions or wordsearches based around the topics, providing places for people to write their own ideas and by giving instructions for the households receiving the zine to do their own waste audit.

The data curation principles and practical activities that are in evidence in the zine are:

1. Create links between data and stories it is used in
2. Simplify data visualization to make it easier to read
3. Try to make data curations as *relatable* to the audience as possible, for example using data from the local environment, or personalizing the data if possible
4. Use storytelling to frame the experiences with the data and guide the points of interaction
5. Give *worked examples* of data inquiries and prompt ways to expand upon this towards own interests
6. Engage people with a data set that they didn't personally collect by encouraging them to collect their own data and add to it.
7. Combine creative with practical activities
8. **News article:** make a news article out of datasets

## 7.2 VEDEN ÄÄRELLÄ

Veden Armoilla is a participatory live game that is being created to engage Finnish students in Lahti with research related to water. As a first step in creating the game experience, an online pilot workshop 'Veden Äärellä' was held to trial some of the arts-based approaches in the context of participatory science storytelling. Thus, a participatory story was created as drama and delivered online, to engage an audience of school students with issues related to a local lake. The participatory story combined pre-recorded videos on different topics related to water, a 'presentation layer' which was set in a live studio where the presenter spoke directly to the students, explained what was happening and introduced the clips and activities and the participation itself, which invited contributions from the audience by asking them to write in the chat, or raise and lower hands in response to questions. The data used as evidence was water data from local lake Vesijärvi. The students learned about phosphorous in the lakes by way of an expert who was 'interviewed' and invited to interpret the data. In this case study, the students did not get to participate with any curated data, but they did themselves *produce* data when participating to the story and this was visualized *live* by using two different coloured water poured into two different chambers according to how many 'yes' or 'no' votes, to show if the audience was more or less positive. Feedback from the audience indicated that they found the *overall experience* to be participatory. In a future iteration the case study will be conducted live and the water data itself will be curated using curator principles.

The data curation principles and practical activities that are in evidence in Veden Äärellä are:

1. Create links between data and stories it is used in
2. Try to make data curations as *relatable* to the audience as possible, for example using data from the local environment, or personalizing the data if possible
3. Use storytelling to frame the experiences with the data and guide the points of interaction
4. Combine creative with practical activities
5. **Roleplay:** have a Q&A with an 'expert' and invite audience questions

## 8 SUMMARY

---

This deliverable presents a number of data curation principles for embedding data as evidence within participatory science stories. It further demonstrates how they have been used in ParCos case study activities up to this point.

## 9 REFERENCES

---

- Bach, B. et al. (2016) 'Telling stories about dynamic networks with graph comics', Conference on Human Factors in Computing Systems - Proceedings, pp. 3670–3682. doi: 10.1145/2858036.2858387.
- Bach, B. et al. (2018) 'Design patterns for data comics', Conference on Human Factors in Computing Systems - Proceedings, 2018-April, pp. 1–12. doi: 10.1145/3173574.3173612.
- Chatman, S., 2021. Story and discourse. Cornell University Press.
- Concannon, S., Rajan, N., Shah, P., Smith, D., Ursu, M. and Hook, J., 2020, April. Brooke leave home: Designing a personalized film to support public engagement with open data. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1-14).
- Gaver, W., Boucher, A., Jarvis, N., Cameron, D., Hauenstein, M., Pennington, S., Bowers, J., Pike, J., Beitra, R. and Ovalle, L., 2016, May. The Datacatcher: batch deployment and documentation of 130 location-aware, mobile devices that put sociopolitically-relevant big data in people's hands: polyphonic interpretation at scale. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 1597-1607).
- Kaninsky, M., Gallacher, S. and Rogers, Y., 2018, June. Confronting People's Fears about Bats: Combining Multi-modal and Environmentally Sensed Data to Promote Curiosity and Discovery. In Proceedings of the 2018 Designing Interactive Systems Conference (pp. 931-943).
- Koesten, L.M., Kacprzak, E., Tennison, J.F. and Simperl, E., 2017, May. The Trials and Tribulations of Working with Structured Data: -a Study on Information Seeking Behaviour. In Proceedings of the 2017 CHI conference on human factors in computing systems (pp. 1277-1289).
- Loukissas, Y.A., 2019. All data are local: Thinking critically in a data-driven society. MIT Press.
- Schorch, P., 2013. Museum encounters and narrative engagements. The International Handbooks of Museum Studies, pp.437-457.
- Wang, Z. et al. (2020) 'Cheat Sheets for Data Visualization Techniques', Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–13. doi: 10.31219/osf.io/z4cxa.

- Wang, Z., Bach, B. and Dingwall, H. (2019) 'Teaching data visualization and storytelling with data comic workshops', Conference on Human Factors in Computing Systems - Proceedings, pp. 1–9. doi: 10.1145/3290607.3299043.
- Wolff, A., Wermelinger, M. and Petre, M., 2019. Exploring design principles for data literacy activities to support children's inquiries from complex data. *International Journal of Human-Computer Studies*, 129, pp.41-54.
- Yan, H., Wang, J. and Xia, C. (2017) 'Research and Application of the Test Data Visualization', Proceedings - 2017 IEEE 2nd International Conference on Data Science in Cyberspace, DSC 2017, pp. 661–665. doi: 10.1109/DSC.2017.110.
- Zhang, J. and Zhao, F. (2017) 'A study on library service innovation based on data curation', Proceedings - 2017 13th International Conference on Semantics, Knowledge and Grids, SKG 2017, 2018-January, pp. 213–218. doi: 10.1109/SKG.2017.00044.

## 10 APPENDIX A – CURATOR CARDS

---

