imec
embracing a better life

WHITE PAPER

WHITE PAPER ON EU BICYCLE DATA SPACES

How data spaces contribute to better and safer cycling experiences

Contributors
Jan Adriaenssens
Stefan Lefever
Evelien Marlier
Bart Matthys
Pieter Morlion
Paul Theyskens
Casper Van Gheluwe

Version
October, 2023

About this white paper

Bicycle Data Spaces (BDS) have emerged to manage and interpret the diverse data generated by the various actors in the cycling ecosystem. As cycling grows in popularity as a mode of transportation, BDS provide an open and interconnected data ecosystem that can help create a more sustainable, efficient, safe, and user-friendly cycling experience.

This paper explores the key concepts, challenges, and opportunities associated with BDS, with a focus on its potential to address the challenges of accessing and interpreting cycling data in the EU. It highlights the benefits of data sharing and collaboration between stakeholders, including governments, bike-sharing companies, and cyclists themselves.

Privacy, security, and standardization are also addressed, with recommendations for addressing these concerns to ensure that the data collected and analyzed is accurate and secure. A Bicycle Data Space offers great opportunities for improving the cycling experience for everyone involved. Finally, this paper provides guidance on how to get started with data spaces effectively.

Contents

I. The rise of cycling, and the need for data

The bicycle is increasingly important for how people get around (Key figures on bikes and cycling in Belgium, FPS, 2023 [1]), partly due to significant efforts by governments and various actors in civil society. This increase has led to a range of implications:

(I) A significant increase in the number of cyclists on the road means that more coordination is needed: between cyclists, between cyclists and cars, between cyclists and pedestrians, and between cyclists and infrastructure.

(II) With the most obvious measures and interventions now (being) implemented in many places, it is key to assess which further interventions could have the largest impact: where is the greatest cycling potential? Which bicycle infrastructure needs urgent maintenance or upgrading? Moreover, once potential improvements have been implemented, measuring their impact is not straightforward.

(III) As bicycle use increases, some negative effects or inappropriate behaviors become more noticeable: bicycle congestion, accidents [2], nuisance to pedestrians, abandoned bikes, haphazard parking, etc.

To fully understand these implications and take the required policy measures, it is necessary to get a more complete picture of cycling behavior: how many people cycle, why they cycle, which routes they take, when they cycle, to what extent infrastructure is used, how effectively new measures are followed up, etc. Real-time data sharing can lead to improved quality, safety, and efficiency for cyclists, as well as to better detection of cyclists (at traffic lights for example).

These questions could largely be answered by better interpretation and exchange of different data sets. However, there is currently not as much data available on cycling as there is on motorized traffic. Nonetheless, cycling data is attracting growing attention (the EU cycling strategy [3] for example). The main problem with its actual use is that these data sets are very scattered, cannot all be interpreted in the same way, and are not always readily accessible.

"As a cyclist, I want to be able to report potholes, faulty lighting, poorly adjusted traffic lights, obstructions, dangerous situations and areas, etc. as easily as possible so that the situation for all cyclists improves as quickly as possible."

(Karine Haerinck, daily bike commuter)

"The Belgian newspaper Het Laatste Nieuws conducted a large-scale survey (VeloVeilig [4]) among cyclists on their perception of safety. It is an enormous amount of work to clean up this data. Both we and the Flemish Cyclists' Association are doing this separately. It would be much more efficient if we could create a good dataset together with other organizations, which in turn could serve as input for further actions and projects."

(Ilse Eylenbosch, Coordinator Bicycle, Agency for Roads and Traffic, Flanders)

II. A specific demand for cycling data

Especially when building new data sets, or making existing data sets interoperable, it is crucial that this data, and the corresponding insights, address the issues experts in the field are struggling with. In the sections below, we investigate questions, problems and barriers outlined by leading actors in the field.

"As a city, we want to have a good understanding of the development of the modal split and bicycle safety so that we can see if our efforts for cyclists are paying off in the long run and whether we are achieving our goals."

(Tim Asperges, Mobility Advisor, City of Leuven)

III. What is an urban mobility data space?

The European Commission is pursuing economic independence and social progress by focusing on data through the European Strategy for Data [5], the European Data Governance Act [6] and the Data Act [7], in which data spaces play a crucial role: to increase data availability, to overcome technical barriers, and especially to give a greater role to data ownership.

Over the past two decades, there have been many laudable attempts to make data freely available through open data portals and interfaces. It takes a lot of time to find, retrieve, interpret, and link data with other datasets. Moreover, it is almost impossible to do this for every city in Europe: information is shared in different formats, the terms of use vary, and even simple definitions, such as what is considered as a bicycle, vary from country to country. This makes it virtually impossible, for example, to link mobility data with data on weather or spatial planning on a European scale. Finally, some work, like data gathering and data cleaning, is often repeated multiple times because processed data and results are not shared again.

A data space can be seen as an evolution of the separate interfaces for (open) data. Such a data space brings together a number of these interfaces in a protected 'virtual space' where very clear agreements and contracts apply. These include how, and with whom, data is exchanged. They furthermore cover concerns such as privacy, security, costs, and intellectual property. Different agreements can also be made between different actors: for example, a city can share data with other governments as free open data, while charging commercial users. By standardizing agreements and ways of working together, it becomes much easier and faster to share or reuse data.

On a technical level, a data space can optimize data interoperability (models, formats, exchange APIs, ...). Some standard building blocks have already been developed by organizations such as the Eclipse Foundation [8] (the Eclipse Dataspace Connector ecosystem) and FIWARE [9] (the TRUE Connector), or private companies such as Sovity [10] (the Eclipse Dataspace user interface). These components allow organizations to share their data in a data space. In addition, there are building blocks that provide identity management, granting of access, security, quality assurance, certification, and transaction traceability. Finally, a data space can facilitate data value creation, such as through data usage accounting and data discovery, and makes it possible to provide extra services such as linking datasets, building visualizations, or storing historical data.

A Flemish, a Belgian and / or a European data space?

Both locally and internationally, many players are working together on the development of data spaces. They are also looking at how specific connectors can be used to build bridges between different ecosystems. Imec therefore collaborates with EDIC and the International Data Space Association (IDSA) and links up with other ongoing EU initiatives like the European Mobility Data Space (EMDS). Imec also participates in the Belgian Data Space Alliance (BDSA) [11], which serves as the local hub of Gaia-X and IDSA. There, several working groups, including one on mobility, will discuss the technical, business and legal aspects of data spaces. In addition, we frequently check the Data Space Radar [12], a website which lists most of the ongoing EU initiatives.

IV. Advantages and disadvantages of a data space for cycling projects

Data spaces can help overcome some of the current hurdles and issues surrounding the use and sharing of data, but they also have drawbacks. This section provides a brief overview and indication of the types of projects for which a data space might be a good solution.

REUSABLE BUILDING BLOCKS

At the basis of data spaces are strong agreements on what the data should look like and how it should be exchanged. The reusable and structured processes, rules and open-source building blocks ensure that information, once part of a data space, can be easily exchanged and combined. For example, when every city or country publishes its bicycle data in the same way in a data space, it is easy to aggregate that data and create a single view of the counts from bike counting poles across Europe. That is what is envisioned with the creation of the Cycle Data Space (CDS) in the MegaBITS project [13] for example. Applications built on top of this data can also be easily reused. For instance, a dashboard built to visualize modal shift in Barcelona should also be perfectly usable in Oslo.

With leading research institutions and the European Commission putting a strong focus on data spaces, it is expected that they will soon become widespread in the EU, and across different sectors. Efforts are not limited to cycling data. In parallel, information on spatial planning, precipitation, air quality, accessibility, trains, etc. is also being published in data spaces. This will make it much easier to share data across policy areas. In addition, the EU aims to create a more level playing field by democratizing access to data, including for smaller players. The focus is on cooperation, not competition.

Furthermore, information will be much easier to find. The data in data spaces will be published in a uniform way in easily searchable 'catalogs'. Thus, all available data in the EU or even in the world could theoretically be searched via a central 'federated' search engine. This also applies to data that has been improved, reworked, or modified. Parties who previously downloaded datasets and improved them locally can now republish them, so that they can be found by everyone. This way, not every data user has to make the same edits, meaning that data users can build on what has already been done. Let's imagine several companies that count passing bicycles provide their data in one consistent format. A separate entity in this data space can then collect all this data. They can combine and process this data from all the companies for a specific area. After that, they can share their findings and automated summaries back in that same data space. Policymakers can then access and potentially pay for these findings to help them make well-informed choices.

CONTROL, QUALITY AND TRACEABILITY

Compared to "traditional" open data, data spaces offer much more control over what happens to the data, which creates more trust in the practice of sharing data. Data can be shared with different parties under different conditions. For example, a citizen science project can share bike counts with governments and cycling organizations for free. For commercial parties, such as a chain of bike stores looking for a new location, a fee can be charged for the same data or conditions can be attached to it. Alternatively, some parties may be able to retrieve detailed data (e.g., the circumstances of each accident in which a cyclist was involved), while others will only have access to aggregate data (such as the number of accidents per postal code).

Quality control also becomes much easier in data spaces. Within a data space, it is possible to implement reusable building blocks that screen data for consistency, continuity, and conformity according to an agreed-upon standard. This can be set up by the publisher of the data or by a third party, such as a commercial or regulatory party.

Moreover, components or parties can be certified within the data space or by the competent authorities to perform certain checks, assign quality labels or anonymize or aggregate data.

In addition, a data space provides standardized agreements, processes, and legal agreements in line with national and European regulations. An application that wants to exchange data can thus reuse existing license agreements, without the need for customization each time. This eases the regulatory burden on ecosystem participants, as the data space is essentially designed to comply and evolve with legislation in areas such as data processing, data management and data protection. Applying these regulations during identity management and data transfer is a major point of focus in the data space domain, resulting in fewer problems with data ownership, data breaches and data misuse.

Finally, traceability plays an important role as well. Within the data space, all transactions go through a 'clearing house' that keeps track of who exchanged what data at what time, which enables settling billable data usage between participants.

STIMULATING THE DATA ECONOMY

Refusing to release information or sharing data on a selective basis by different public or private organizations stifles innovation in the data economies of the EU and its member states. Providing controlled and equal access to data gives future innovations, such as in the field of AI, a fair chance to be rolled out across Europe and offers all players an equal opportunity to develop applications on top of the data. It also enables smaller players to become part of new and existing ecosystems and allows new business models to merge.

In terms of business models, a data space opens a world of possibilities. Because a data space enables a hybrid model, a data owner no longer must choose between open data and paid licenses. Different business and licensing models can be linked to the same data source, depending on the customer, the intended use, whether the data needs to be made available for reuse, etc. Moreover, when operations and analysis based on the data are reusable, they can be re-offered within the data space - either as a paid service or not.

"As a crowdsensing platform and community, we want to continue publishing open data while monetizing commercial use so that we can cover our operating costs."

(Kris Vanherle, Coordinator, Telraam)

Through standardization of technical, legal, and business agreements, along with reusable components and a clear division of tasks and responsibilities, turnaround times are greatly reduced. This is true both for publishing data and for applications that reuse the data. Once the data space has been set up, there can be significant time-to-market advantages over traditional data sharing, IT integration and processing projects.

Considerations when publishing data in a data space

Data spaces solve many shortcomings of current data ecosystems, but they do not relieve data publishers of all responsibilities. Before onboarding data onto a dataspace, the data must meet minimum data quality requirements: it needs to be formatted, standardized and linked. In other words, a data space does provide standard building blocks and a number of requirements for any data to be onboarded. But it does not define data standards. Parties must first agree how they will format the data. Specifically for cycling data, there are few data standards at a European level. For example, bicycle navigation apps typically use the GPS Exchange Format (GPX) [14]. To exchange bicycle counts, you need a common standard that is used across the EU. Different standards such as TOMP or MDS also exist for shared bicycles. The advantage of a data space is that data can easily be converted or shared in multiple formats if the same definitions are used.

One common misconception about publishing data in a data space is that this requires a central data repository that stores the shared data for a long time. But in fact, the complete opposite is true. The data always remains with the data providers. Only after a consumer and provider reach a data transfer agreement, the data, or portion of the data, that they agreed to share is exchanged automatically and directly from the provider to the consumer, without the need of an intermediary.

In the long term, data spaces can become a very powerful way to share data and achieve great efficiencies in terms of interoperability and reuse, as well as increasing privacy, trust and data security. Even in the short term, these benefits are already there, but they do not apply to every environment or every type of problem. Setting up a data space is an intensive exercise that requires a lot of coordination and often difficult agreements between many actors, not to mention the issues of technical implementation. In theory, a data space is a cost-effective solution, especially for intensive or complex data exchanges between private and public parties. However, for simple projects where data needs to be shared between a limited number of actors or for short-term solutions, the costs may be disproportional to the benefits. It is therefore advisable to begin by gathering the specific needs of stakeholders in the ecosystem and analyze which use cases will have the greatest impact and require sustainable long-term implementation.

V. What types of projects can a data space be useful for?

There are several hurdles to overcome before a large, pan-European data space becomes a reality. Perhaps, such a data space will be created through a series of small steps that need to be taken each time a new project is set up, new regulations are introduced, and existing systems are renewed. It is advisable to start with projects with a specific immediate application, without losing sight of the long-term vision and architecture and by connecting to other initiatives, so that the convergence of these small projects into federated European data spaces becomes an obvious course of action.

Subdomains of cycling data/data spaces

Strategic research center imec took the lead in Belgium, bringing together the most committed cycling organizations and governments at the end of 2022 to consider the following questions: "What data do you already have?", "What data do you still need?" and "What would you like to use that data for?" The exercise resulted in a useful overview [15] of the ongoing bicycle data projects. In addition, a roadmap [16] was made of all projects in the pipeline and will be implemented soon, so that synergies between different stakeholders can be sought.

"As a province, we want to be able to properly assess the cycling potential of certain locations and interventions. There are about a hundred places in our provinces where we could build infrastructure such as a bridge, but we don't have sufficient insight into which intervention would have the greatest positive impact."

(Anke Schelfaut, Department Head Mobility, Province of East Flanders)

The Belgian projects and needs have been mapped, applying different subdomains of cycling data (e.g. trip data, parking data, counting data). Below is an overview of how a data space could support specific needs and projects in Belgium.

"Road authorities want to know where cyclists are, to enhance their traffic safety. Intelligent traffic lights, for example, create safer and more comfortable flows for cyclists when they give extra green time. We want to service cyclists with a personalized approach at intelligent traffic lights and real-time notifications about potentially dangerous situations on their way."

(Ben Helsen, Director-Engineer Innovation, Agency for Roads and Traffic, Flanders)

1. **Point data: how many cyclists pass a certain point?**
   There are numerous bicycle counting systems and campaigns, from bicycle counting poles to traffic studies and citizen initiatives. This data is currently not aggregated, making it impossible to conduct area-wide analyses. However, a data space where commercial use of the data must be paid for could, for example, provide some citizen and crowdsource initiatives with the funding they need to keep their systems and operations running.

   Standardization within a data space can also ensure that it is very easy to add new counts to an overarching dashboard or analysis tool.

2. **Data on movement and route planning: where does the cyclist ride or want to ride?**
   While there are already some counts at certain points, there is still little information about the origin and destination of the cyclist, let alone the exact route he or she took. That data is nevertheless remarkably interesting for analyzing the use of existing infrastructure and future cycling potential, planning new infrastructure, prioritizing bicycle routes, and making safety analyses. A data space could provide the components and analysis tools so that small-scale campaigns can bring the data together without having to set up infrastructure and data processing each time. Or ideally, the data space could provide a system where individual users can share the data from their cycling apps or phones with their municipality to drive cycling policy development.

3. **Parking data: where are bicycle parking facilities and how much are they used?**
   Belgium already has a centralized and linked open data register of bicycle parking facilities. Apps are also linked to it, for example to set up campaigns to count the number of bicycles in parking facilities. By working with the standardized components, a data space could easily be further linked to other applications and other bike services from other countries.

4. **Data on shared bikes: where are shared bikes located and how many are available?**
   There are already a few standards and exchange mechanisms in the field of shared bikes. However, these still require some implementation work and are not interoperable. Offering related components in a data space could improve this. Governments often require operators to share data, but do not always have the resources to process it. If operators were to publish the data immediately in a data space, the data could be added in a standardized format and meet the minimum data requirements, which would save governments a lot of extra time and effort.

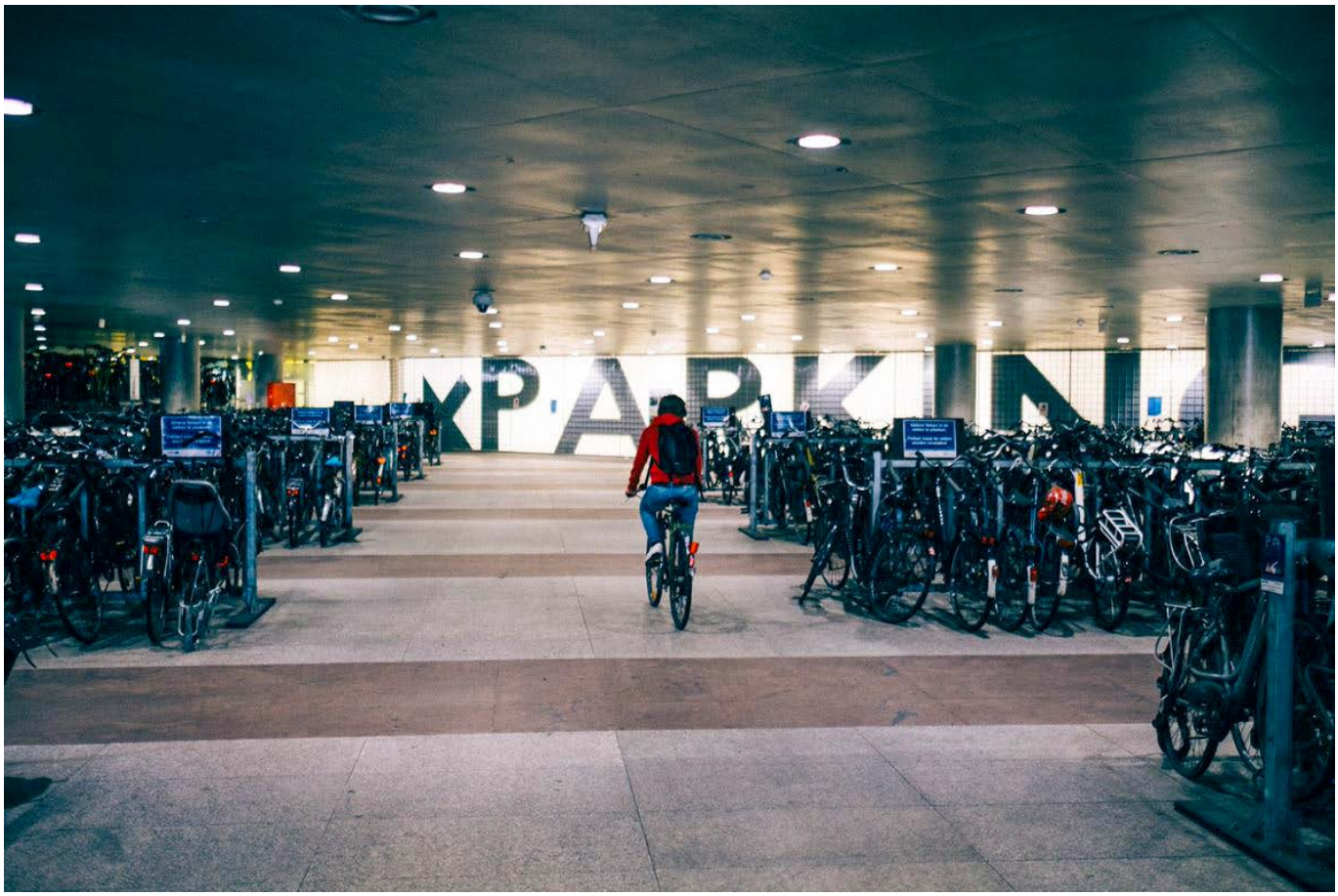
5. **Data on assets: information about bike lanes, bike paths and infrastructure.**
   Up-to-date and reliable information about bike lanes, bike paths and other bicycle infrastructure (e.g., bridges, underpasses, parking lots, bicycle pumps, and by extension traffic signs) is fundamental to any project on bicycle data. Only when the digital version of reality is sufficiently accurate, asset management can be optimized, and missing links be discovered. And most importantly, only when the cycling infrastructure is made available digitally, route planners can suggest the best and safest cycling route. Currently, there are many maps of bike lanes, but they are often not interchangeable. By publishing them in a standardized format in a data space, the data could be more easily updated, shared, and used by cycling applications such as navigation tools.

6. **Data on route experience: information about journey quality and experience.**
   There are several projects that use a bicycle equipped with sensors to map bike lanes and their quality. However, this data is not shared in an optimal way. In addition, there are a few projects that collect input from cyclists and bicycle organizations, for example, on how a cyclist experiences a particular route or place. A data space could ensure that this data can be brought together, and that a holistic view could be created for a particular area or type of report.

7. **Data on road safety: where are dangerous and potentially dangerous points located?**
   Many of the governments and organizations surveyed attach great importance to reducing the number of accidents involving cyclists. However, there is rarely a good and detailed overview of where accidents with cyclists occur, so-called 'black spots', or of their causes; let alone a good understanding of the definition of a dangerous place and where accidents could potentially happen ('gray spots'). In a data space, existing objective data could be merged with subjective data and reports from cyclists, and linked to other information (e.g., weather, speed, bike lane width).

8. **Data on the bicycle itself: characteristics and identification.**
   For several decades, some bicycles have been registered using a national register number engraved on the frame, followed by newer systems using stickers. Systems that store centralized data on bicycles could publish information on individual bicycles (e.g., frame number, photo, color, features, owner) in a data space so that trusted links could be made with other systems for maintenance, resale, or theft reporting.

9. **Data on cyclists: who is cycling?**
   Cyclists are regularly surveyed by various parties and at various levels (e.g. local, regional). Unfortunately, the results of the surveys are often difficult to exchange. Making good arrangements and publishing the data in a data space can ensure that all data is available in the same analysis tool so that a more complete picture can be formed of the needs and wishes of cyclists.



VI. Publishing data on cycling in a data space: how to get started?

Organizations that want to get started with data spaces can contact the Data Spaces Support Centre [17], among others. They offer support and on their website you can find a data space starter kit [18], alongside best practices and advice on which components can be implemented. Gaia-X also offers a starter document [19] and IDSA gives guidelines on implementing a data space based on their reference architecture [20]. Sitra, the Finnish Innovation Fund, also organizes regular webinars on data spaces including one on "setting up data spaces as a service" [21].

In terms of funding, there are several ways to get support, mainly from the European Commission. For example, it offers Cascade Funding [22], also called Financial Support to Third Parties. This is a mechanism where support is provided on behalf of the European Commission, for example through DS4SSCC [23]: the Data Space for Smart and Sustainable Cities and Communities initiative. Funding for data spaces is also provided by ERDF [24] and Interreg [25].

Governments and organizations do not need to publish their data in a data space or build and maintain all the components themselves. In fact, data spaces make it easier to outsource certain tasks and keep the focus on core activities. An interesting possibility is that organizations which purchase certain supplies or services (e.g., bike counting infrastructure or the mapping of bike lane quality) could require the supplier to publish the information in a data space as part of the contract. In this way, new applications would systematically become part of a data space, creating an ecosystem in which it becomes increasingly interesting and easy for all members to develop new applications.

Imec Data Space Competence Center

As a strategic research institution, imec wants to share the knowledge we have accumulated with others as much as possible – including our expertise on data spaces. Do you have a question about **data spaces**? Would you like to test with a neutral party whether a data space can be an added value for your use case/project? Then contact **bart.matthys@imec.be**.

Do you have questions about cycling data or about this paper in particular? Feel free to contact **evelien.marlier@imec.be**.

References

[1] **Kerncijfers van de fiets in België:** https://mobilit.belgium.be/sites/default/files/documents/publications/2023/Kerncijfers%20van%20de%20fiets_NL.pdf

[2] **Verkeersveiligheidsbarometer:** F6. Fietsers https://www.vias.be/storage/main/barometer-2022-1-nl.html#F6_Fietsers

[3] **EU Cycling Strategy:** Recommendations for Delivering Green Growth and an Effective Mobility System in 2030 https://ecf.com/system/files/EUCS_full_doc_small_file_0.pdf

[4] **VeloVeilig:** https://www.hln.be/dossier/veloveilig-d081970f1-b55e-4a39-9732-54f1c78d25cb/

[5] **A European Strategy for data:** https://digital-strategy.ec.europa.eu/en/policies/strategy-data#:~:text=Common%20European%20data%20spaces%20will,improve%20health%20care

[6] **European Data Governance Act:** https://digital-strategy.ec.europa.eu/en/policies/data-governance-act

[7] **Data Act:** https://digital-strategy.ec.europa.eu/en/policies/data-act

[8] **Eclipse Dataspace Components:** https://projects.eclipse.org/projects/technology.edc

[9] **FIWARE TRUE Connector:** https://fiware-true-connector.readthedocs.io/en/latest/

[10] **Sovity:** https://github.com/orgs/sovity/repositories

[11] **Belgian Data Spaces Alliance:** https://www.dataspacesalliance.be/

[12] **Data Space Radar:** https://internationaldataspaces.org/adopt/data-space-radar/

[13] **MegaBITS:** https://www.interregnorthsea.eu/megabits/about-us

[14] **GPS Exchange Format:** https://en.wikipedia.org/wiki/GPS_Exchange_Format#:~:text=GPX%2C%20or%20GPS%20Exchange%20Format,need%20to%20pay%20license%20fees.

[15] **Overzicht fietsdata:** https://docs.google.com/document/d/1-xavC2Tn159hPGqamQyOqaOQOkO9WDuBl1jibAPp_0k/edit

[16] **Fietsdata:** overzicht projecten en noden 2023 – 2024 in Vlaanderen / België https://docs.google.com/document/d/1mUtFrZkfTwON6-uMSleEEcNaTJMMC3XhBcV61AD0P2o/edit

[17] **Data Spaces Support Centre:** https://dssc.eu/

[18] **Starter Kit for Data Space Designers:** https://dssc.eu/space/SK/29523973/Starter+Kit+for+Data+Space+Designers+%7C+Version+1.0+%7C+March+2023

[19] **A practical user perspective on building data spaces:** https://onedrive.live.com/?authkey=%21ACrUFUeTq2dFgmk&id=FCF5D22AF37A710C%2144681&cid=FCF5D22AF37A710C&parId=root&parQt=sharedby&o=OneUp

[20] **IDS Reference Architecture Model:** https://internationaldataspaces.org/publications/ids-ram/

[21] **Setting Up Data Spaces as a Service:** https://www.sitra.fi/en/events/data-spaces-as-a-service/

[22] **Cascade Funding Calls / Financial Support for Third Parties (FSTP):** https://webgate.ec.europa.eu/funding-tenders-opportunities/pages/viewpage.action?pageId=25559615

[23] **Data Space for Smart and Sustainable Cities and Communities:** https://www.ds4sscc.eu/

[24] **European Regional Development Fund:** https://ec.europa.eu/regional_policy/funding/erdf_en

[25] **Interreg Europe:** https://www.interregeurope.eu/