

Norwegian Scientific Data Network

Øystein Godøy
Bard Saadatnejad



Background

- National e-infrastructure
- Funded by Research Council of Norway
- 5 year project
- 10 year initial time line
 - i.e. a commitment to continue operation beyond development
- 5 partners
 - METNO, IMR, NPI, NERSC, NILU



Background

- National e-infrastructure
- **Funded by Research Council of Norway**
- 5 year project
- 10 year initial time line
 - i.e. a commitment to continue operation beyond development
- 5 partners
 - METNO, IMR, NPI, NERSC, NILU



Background

- National e-infrastructure
- Funded by Research Council of Norway
- **5 year project**
- 10 year initial time line
 - i.e. a commitment to continue operation beyond development
- 5 partners
 - METNO, IMR, NPI, NERSC, NILU



Background

- National e-infrastructure
- Funded by Research Council of Norway
- 5 year project
- 10 year initial time line
 - i.e. a commitment to continue operation beyond development
- 5 partners
 - METNO, IMR, NPI, NERSC, NILU



Background

- National e-infrastructure
- Funded by Research Council of Norway
- 5 year project
- 10 year initial time line
 - i.e. a commitment to continue operation beyond development
- **5 partners**
 - METNO - Norwegian Meteorological Institute
 - NILU - Norwegian Institute for Air Research
 - IMR - Institute of Marine Research
 - NPI - Norwegian Polar Institute
 - NERSC - Nansen Environmental and Remote Sensing Center



Purpose

- The vision
 - Safe stewardship, preservation and unified access to heterogeneous scientific datasets in the environmental domain.
- This project aims at providing the scientific community with an integrated, cost-efficient and sustainable infrastructure following established standards for data documentation, archiving, search and exchange.
- The intention is to link existing institutional and discipline specific systems to promote science regardless of geographical and institutional location.



Purpose

- The vision
 - Safe stewardship, preservation and unified access to heterogeneous scientific datasets in the environmental domain.
- This project aims at providing the scientific community with an integrated, cost-efficient and sustainable infrastructure following established standards for data documentation, archiving, search and exchange.
- The intention is to link existing institutional and discipline specific systems to promote science regardless of geographical and institutional location.



Purpose

- The vision
 - Safe stewardship, preservation and unified access to heterogeneous scientific datasets in the environmental domain.
- This project aims at providing the scientific community with an integrated, cost-efficient and sustainable infrastructure following established standards for data documentation, archiving, search and exchange.
- The intention is to link existing institutional and discipline specific systems to promote science regardless of geographical and institutional location.



Objectives (1)

1. Development of a sustainable governance structure

- i. Promote and support interdisciplinary Metadata (Arctic Data Committee) - also outside of the Polar Regions - to link scientists across disciplines and nations.
- ii. Promote international accreditation of the participating data centres.
- iii. Assess funding models to ensure a sustainable interdisciplinary scientific infrastructure.
- iv. Coordinate requests for NorStore functionality from the climate and environmental data community.



Objectives (1)

1. Development of a sustainable governance structure

- i. Promote and support the interdisciplinary ADC Polar Metadata Profile universally – also outside of the Polar Regions – to link scientists across disciplines and nations.
- ii. Promote international accreditation of the participating data centres.
- iii. Assess funding models to ensure a sustainable interdisciplinary scientific infrastructure.
- iv. Coordinate requests for NorStore functionality from the climate and environmental data community.



Objectives (1)

1. Development of a sustainable governance structure

- i. Promote and support the interdisciplinary ADC Polar Metadata Profile universally – also outside of the Polar Regions – to link scientists across disciplines and nations.
- ii. Promote international accreditation of the participating data centres.
- iii. Assess funding models to ensure a sustainable interdisciplinary scientific infrastructure.
- iv. Coordinate requests for NorStore functionality from the climate and environmental data community.



Objectives (1)

1. Development of a sustainable governance structure

- i. Promote and support the interdisciplinary ADC Polar Metadata Profile universally – also outside of the Polar Regions – to link scientists across disciplines and nations.
- ii. Promote international accreditation of the participating data centres.
- iii. Assess funding models to ensure a sustainable interdisciplinary scientific infrastructure.
- iv. Coordinate requests for NorStore functionality from the climate and environmental data community.



Objectives (2)

2. Implementation and maintenance of **distributed heterogeneous** data management
 - i. **Maintain the infrastructure for data management developed during IPY** to support projects and scientists and renew it towards international data management efforts, emphasising the **semantic brokering required to promote interdisciplinary science**.
 - ii. Extend the data management infrastructure from the IPY to more catalogues hosted by the participating institutions in order to establish a unified catalogue.
 - iii. Maintain, improve and further standardize data set documentation to help data filtering and long term stewardship of heterogeneous datasets.
 - iv. Development of tools for data providers simplifying data documentation and submission, with emphasis on discovery and use metadata generation using international standards.
 - v. Standardise the internal interfaces between data centres in order to effectively support a cost effective unified view of process oriented data.
 - vi. Provide direct, online access to datasets, in accordance with the usage or distribution constraints described in the metadata of each dataset.
 - vii. Provide users with best practises, guides and tools for proper use of data and information infrastructures.



Objectives (2)

2. Implementation and maintenance of **distributed heterogeneous** data management
 - i. Maintain the infrastructure for data management developed during IPY to support projects and scientists and renew it towards international data management efforts, emphasising the semantic brokering required to promote interdisciplinary science.
 - ii. **Extend** the data management infrastructure from the **IPY to more catalogues** hosted by the participating institutions in order **to establish a unified catalogue**.
 - iii. Maintain, improve and further standardize data set documentation to help data filtering and long term stewardship of heterogeneous datasets.
 - iv. Development of tools for data providers simplifying data documentation and submission, with emphasis on discovery and use metadata generation using international standards.
 - v. Standardise the internal interfaces between data centres in order to effectively support a cost effective unified view of process oriented data.
 - vi. Provide direct, online access to datasets, in accordance with the usage or distribution constraints described in the metadata of each dataset.
 - vii. Provide users with best practises, guides and tools for proper use of data and information infrastructures.



Objectives (2)

2. Implementation and maintenance of distributed heterogeneous data management
 - i. Maintain the infrastructure for data management developed during IPY to support projects and scientists and renew it towards international data management efforts, emphasising the semantic brokering required to promote interdisciplinary science.
 - ii. Extend the data management infrastructure from the IPY to more catalogues hosted by the participating institutions in order to establish a unified catalogue.
 - iii. Maintain, improve and **further standardize data set documentation** to help data filtering and long term stewardship of heterogeneous datasets.
 - iv. Development of tools for data providers simplifying data documentation and submission, with emphasis on discovery and use metadata generation using international standards.
 - v. Standardise the internal interfaces between data centres in order to effectively support a cost effective unified view of process oriented data.
 - vi. Provide direct, online access to datasets, in accordance with the usage or distribution constraints described in the metadata of each dataset.
 - vii. Provide users with best practises, guides and tools for proper use of data and information infrastructures.

Objectives (2)

2. Implementation and maintenance of distributed heterogeneous data management
 - i. Maintain the infrastructure for data management developed during IPY to support projects and scientists and renew it towards international data management efforts, emphasising the semantic brokering required to promote interdisciplinary science.
 - ii. Extend the data management infrastructure from the IPY to more catalogues hosted by the participating institutions in order to establish a unified catalogue.
 - iii. Maintain, improve and further standardize data set documentation to help data filtering and long term stewardship of heterogeneous datasets.
 - iv. Development of **tools for data providers** simplifying data **documentation** and **submission**, with emphasis on discovery and use metadata generation using international standards.
 - v. Standardise the internal interfaces between data centres in order to effectively support a cost effective unified view of process oriented data.
 - vi. Provide direct, online access to datasets, in accordance with the usage or distribution constraints described in the metadata of each dataset.
 - vii. Provide users with best practises, guides and tools for proper use of data and information infrastructures.

Objectives (2)

2. Implementation and maintenance of distributed heterogeneous data management
 - i. Maintain the infrastructure for data management developed during IPY to support projects and scientists and renew it towards international data management efforts, emphasising the semantic brokering required to promote interdisciplinary science.
 - ii. Extend the data management infrastructure from the IPY to more catalogues hosted by the participating institutions in order to establish a unified catalogue.
 - iii. Maintain, improve and further standardize data set documentation to help data filtering and long term stewardship of heterogeneous datasets.
 - iv. Development of tools for data providers simplifying data documentation and submission, with emphasis on discovery and use metadata generation using international standards.
 - v. **Standardise** the internal **interfaces between data centres** in order to effectively support a cost effective unified view of data.
 - vi. Provide direct, online access to datasets, in accordance with the usage or distribution constraints described in the metadata of each dataset.
 - vii. Provide users with best practises, guides and tools for proper use of data and information infrastructures.

Objectives (2)

2. Implementation and maintenance of distributed heterogeneous data management
 - i. Maintain the infrastructure for data management developed during IPY to support projects and scientists and renew it towards international data management efforts, emphasising the semantic brokering required to promote interdisciplinary science.
 - ii. Extend the data management infrastructure from the IPY to more catalogues hosted by the participating institutions in order to establish a unified catalogue.
 - iii. Maintain, improve and further standardize data set documentation to help data filtering and long term stewardship of heterogeneous datasets.
 - iv. Development of tools for data providers simplifying data documentation and submission, with emphasis on discovery and use metadata generation using international standards.
 - v. Standardise the internal interfaces between data centres in order to effectively support a cost effective unified view of process oriented data.
 - vi. Provide direct, online access to datasets, in accordance with the usage or distribution constraints described in the metadata of each dataset.
 - vii. Provide users with best practises, guides and tools for proper use of data and information infrastructures.

Objectives (2)

2. Implementation and maintenance of distributed heterogeneous data management
 - i. Maintain the infrastructure for data management developed during IPY to support projects and scientists and renew it towards international data management efforts, emphasising the semantic brokering required to promote interdisciplinary science.
 - ii. Extend the data management infrastructure from the IPY to more catalogues hosted by the participating institutions in order to establish a unified catalogue.
 - iii. Maintain, improve and further standardize data set documentation to help data filtering and long term stewardship of heterogeneous datasets.
 - iv. Development of tools for data providers simplifying data documentation and submission, with emphasis on discovery and use metadata generation using international standards.
 - v. Standardise the internal interfaces between data centres in order to effectively support a cost effective unified view of process oriented data.
 - vi. Provide direct, online access to datasets, in accordance with the usage or distribution constraints described in the metadata of each dataset.
 - vii. Provide users with best practises, guides and tools for proper use of data and information infrastructures.

Objectives (3)

3. Long term preservation of datasets

- i. Provide **permanent, unified access to preserved datasets**. (Actual data storage must be handled by existing and emerging systems external to this project, at the host institutions and elsewhere.)
- ii. Development of data preservation guidelines for data centres and data producers (scientists), including usage of unique, persistent and citable identifiers.
- iii. Identify potential higher order services that are useful and needed by scientists to minimise technical obstacles encountered in scientific work.
- iv. Develop and implement higher order services.

Objectives (3)

3. Long term preservation of datasets

- i. Provide permanent, unified access to preserved datasets. (Actual data storage must be handled by existing and emerging systems external to this project, at the host institutions and elsewhere.)
- ii. Development of data preservation guidelines for data centres and data producers (scientists), including usage of **unique, persistent and citable identifiers**.
- iii. Identify potential higher order services that are useful and needed by scientists to minimise technical obstacles encountered in scientific work.
- iv. Develop and implement higher order services.



Objectives (3)

3. Long term preservation of datasets

- i. Provide permanent, unified access to preserved datasets. (Actual data storage must be handled by existing and emerging systems external to this project, at the host institutions and elsewhere.)
- ii. Development of data preservation guidelines for data centres and data producers (scientists), including usage of unique, persistent and citable identifiers.
- iii. Identify potential **higher order services** that are useful and needed by scientists to minimise technical obstacles encountered in scientific work.
- iv. Develop and implement higher order services.

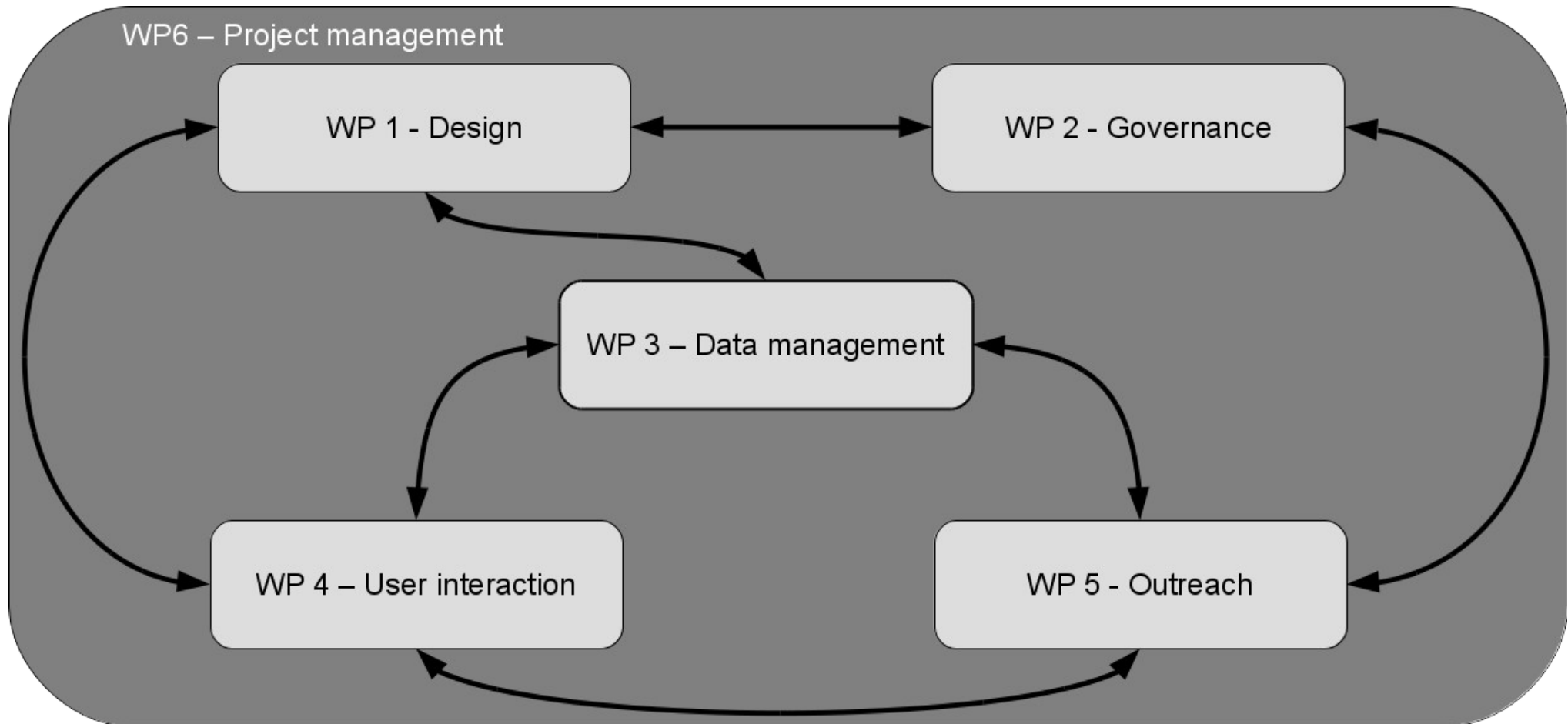
Objectives (3)

3. Long term preservation of datasets

- i. Provide permanent, unified access to preserved datasets. (Actual data storage must be handled by existing and emerging systems external to this project, at the host institutions and elsewhere.)
- ii. Development of data preservation guidelines for data centres and data producers (scientists), including usage of unique, persistent and citable identifiers.
- iii. Identify potential higher order services that are useful and needed by scientists to minimise technical obstacles encountered in scientific work.
- iv. Develop and implement higher order services.

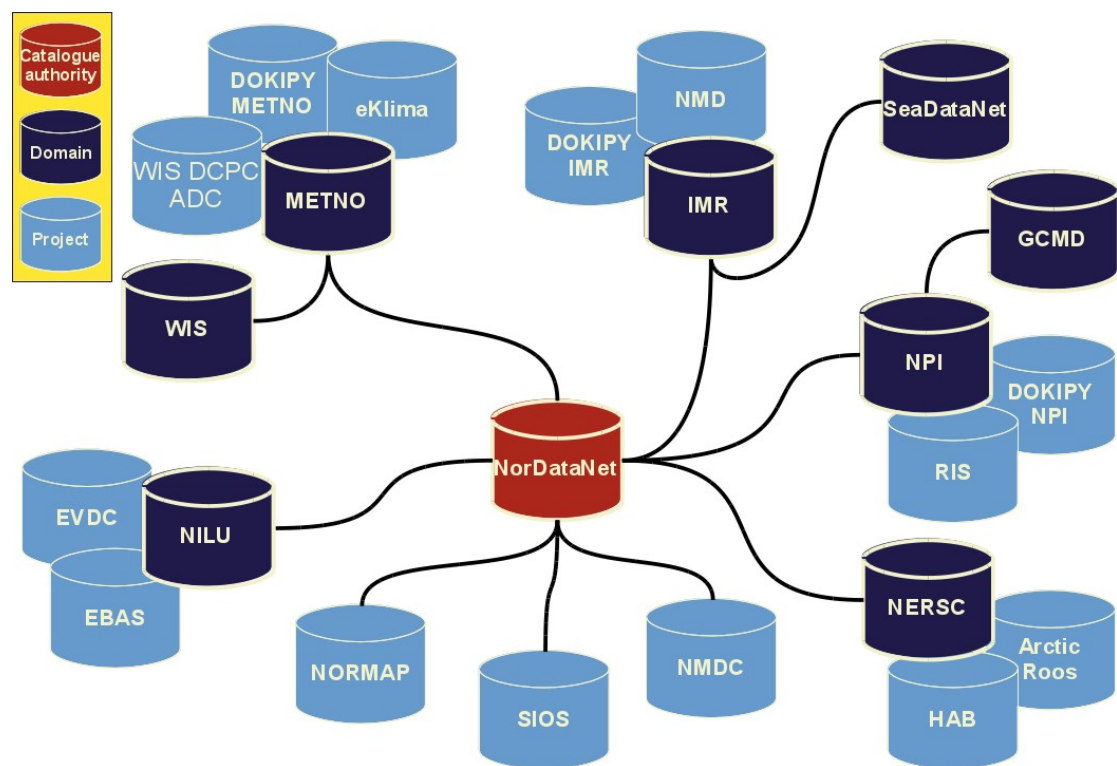


Work packages



The infrastructure (1)

The Norwegian Scientific Data Network (NorDataNet) is a **distributed infrastructure** where existing systems are linked using index metadata to constitute a virtual infrastructure.



The infrastructure (2)

- This project is intended to link infrastructure projects like NorStore, NORMAP and NMDC with existing mandated archives at the participating institutions using internationally accepted standards.
- Further intentions are to create two way linkages with corresponding international facilities and to establish a unified interface that allows scientists to search a number of catalogues simultaneously without having to access each individual portal. This approach will provide global visibility to Norwegian datasets and scientists.

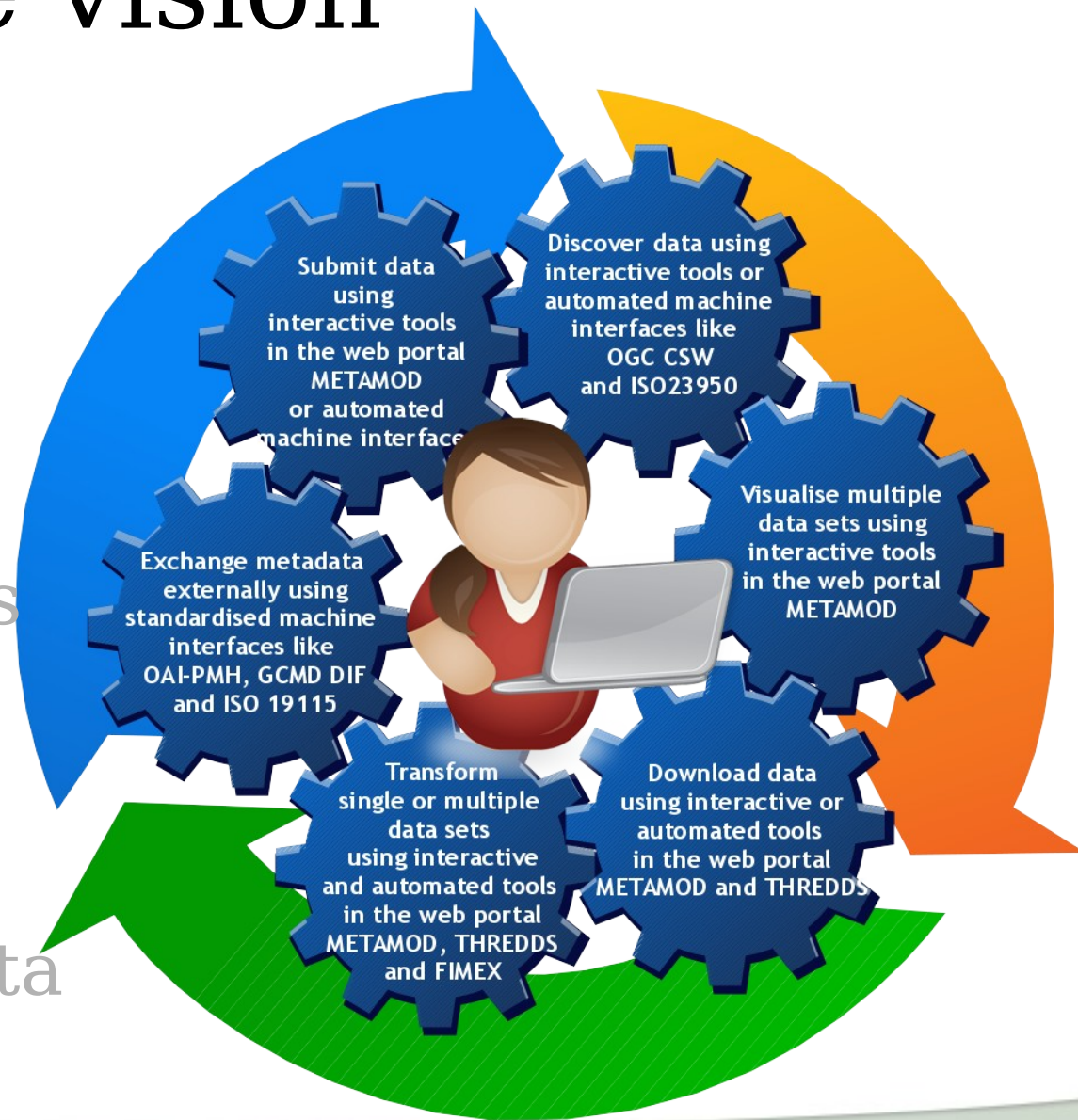


The infrastructure (2)

- This project is intended to link infrastructure projects like NorStore, NORMAP and NMDC with existing mandated archives at the participating institutions using internationally accepted standards.
- Further intentions are to create two way linkages with corresponding international facilities and to establish a **unified interface** that allows scientists to search **a number of catalogues simultaneously** without having to access each individual portal. This approach will provide global visibility to Norwegian datasets and scientists.

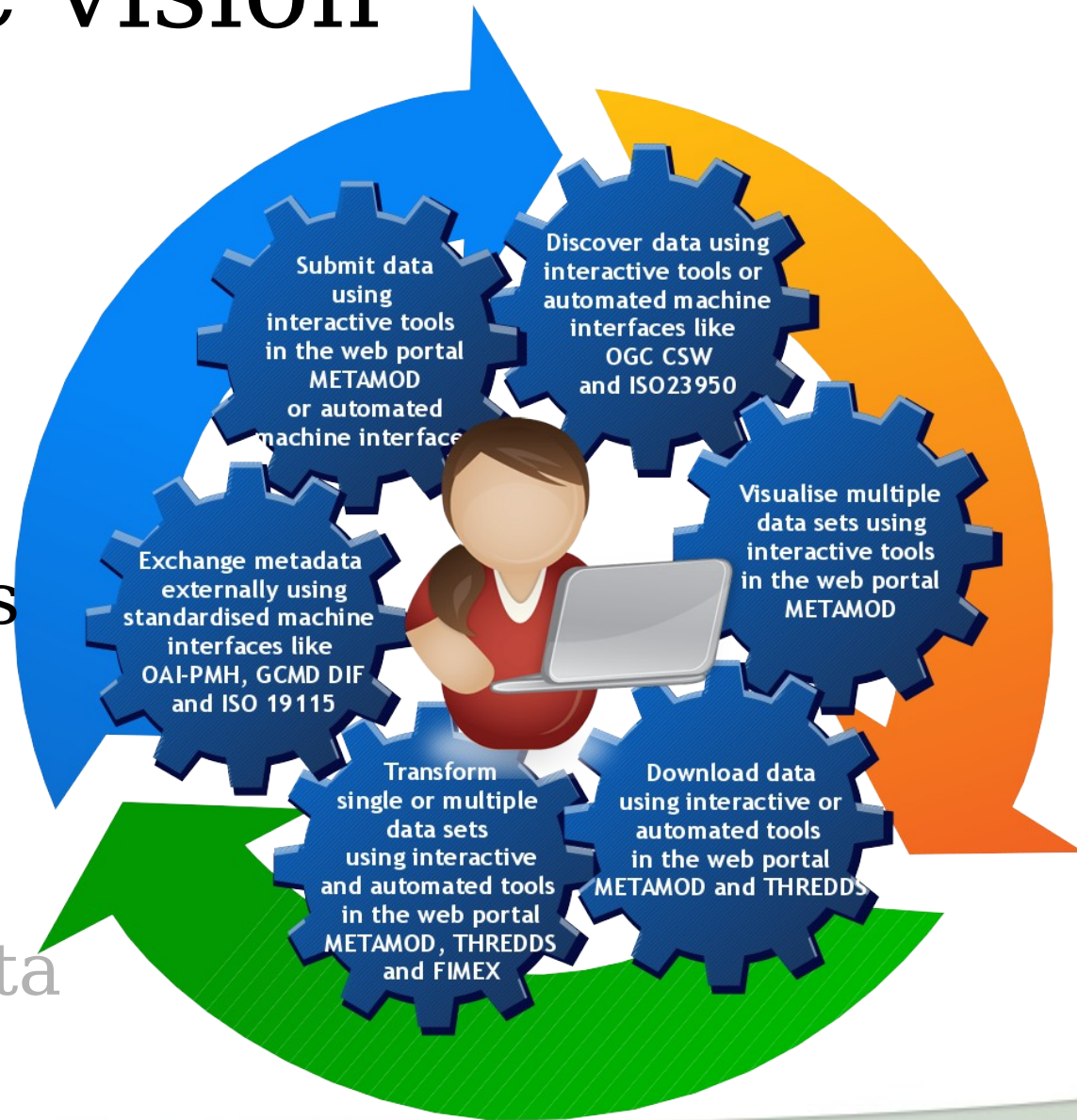
The vision

- A technical toolbox for distributed data management
- For continuous and ad hoc data streams
- Utilisation of internationally accepted interfaces to metadata and data



The vision

- A technical toolbox for distributed data management
- For continuous and ad hoc data streams
- Utilisation of internationally accepted interfaces to metadata and data



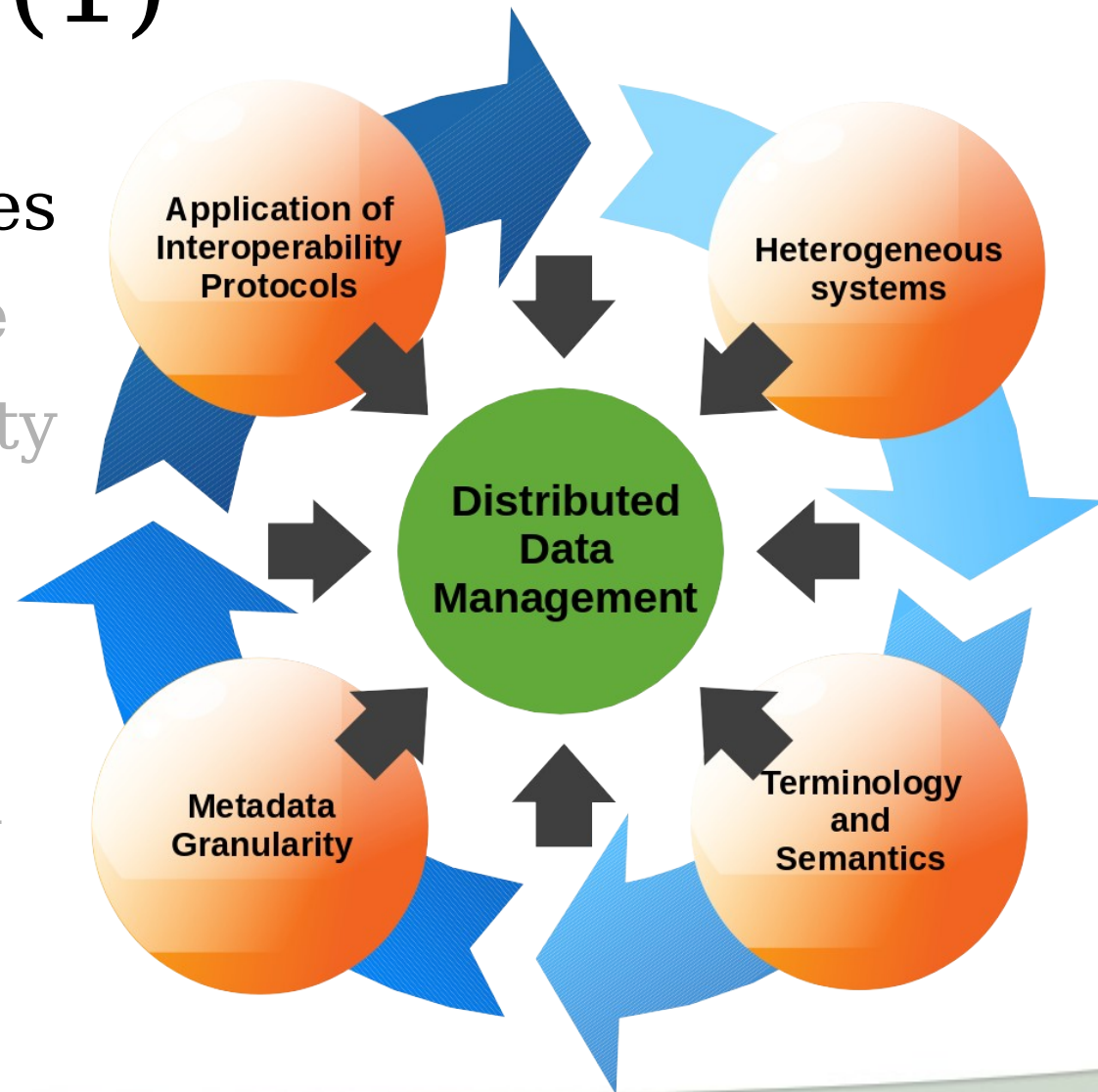
The vision

- A technical toolbox for distributed data management
- For continuous and ad hoc data streams
- Utilisation of internationally accepted interfaces to metadata and data



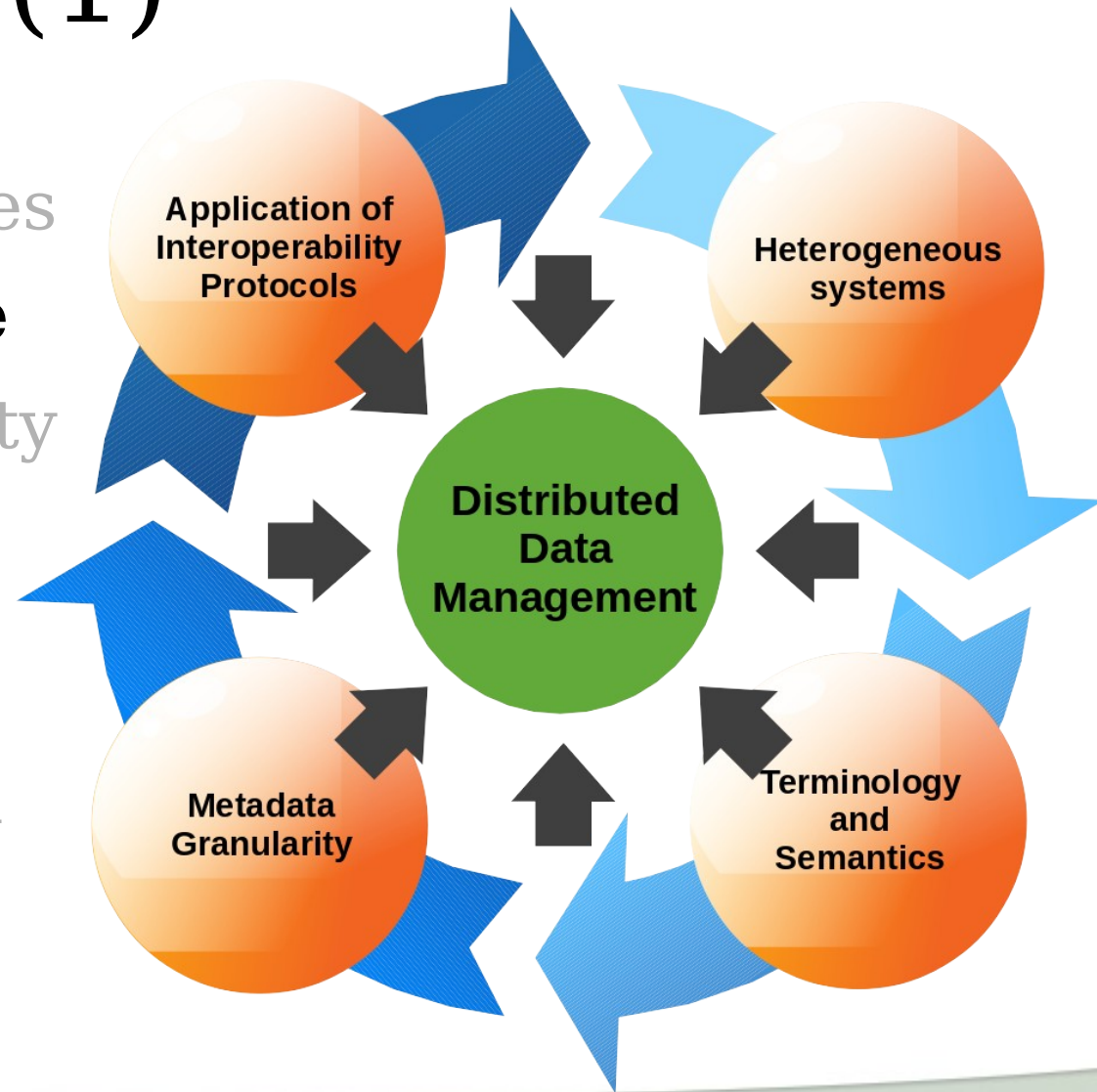
Distributed data management (1)

- Connects physically distributed data centres
- A “virtual” data centre
- Require interoperability
 - Metadata
 - Data
- Powered through documentation of data
 - Discovery metadata
 - Use metadata



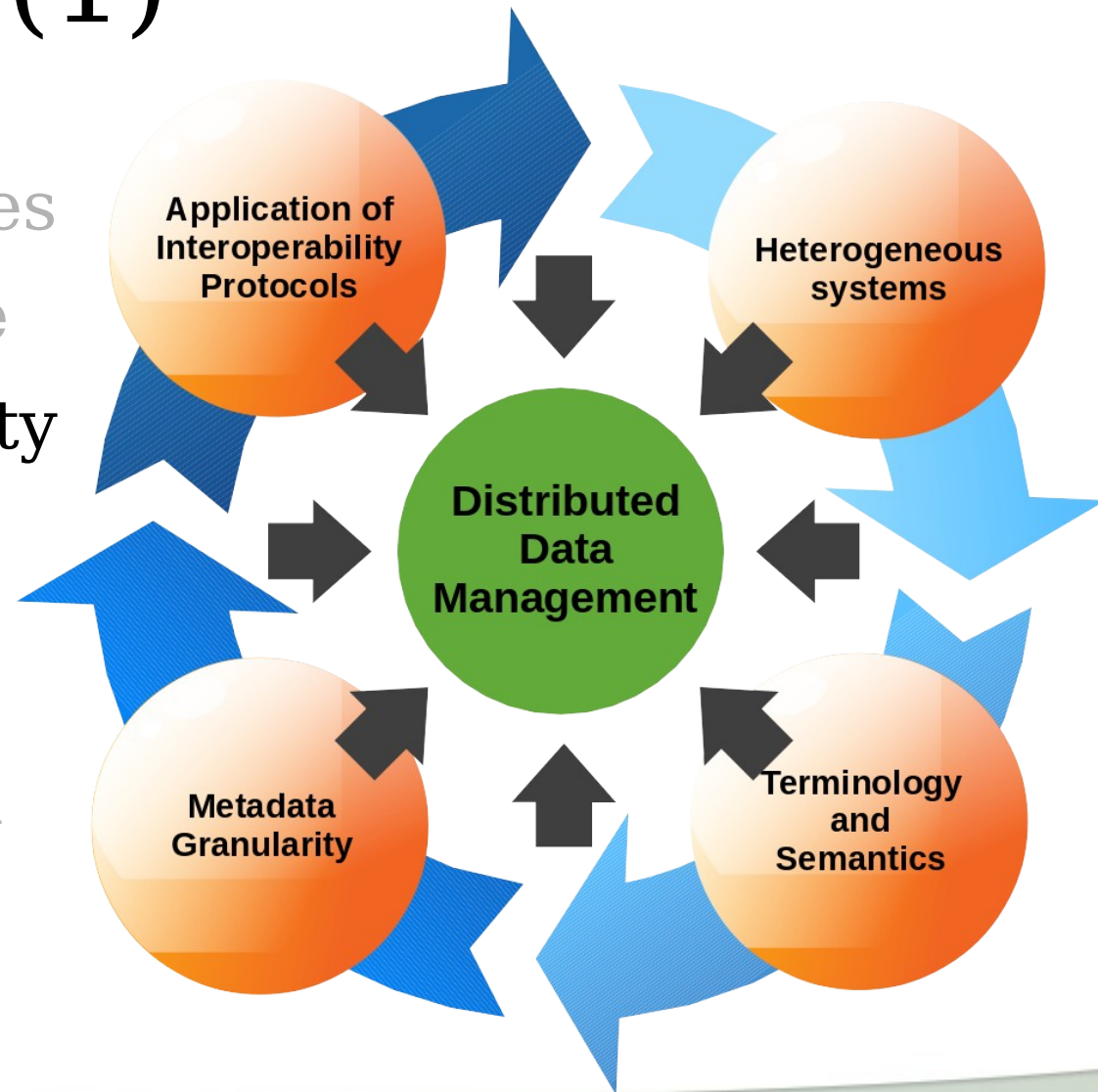
Distributed data management (1)

- Connects physically distributed data centres
- A “virtual” data centre
- Require interoperability
 - Metadata
 - Data
- Powered through documentation of data
 - Discovery metadata
 - Use metadata



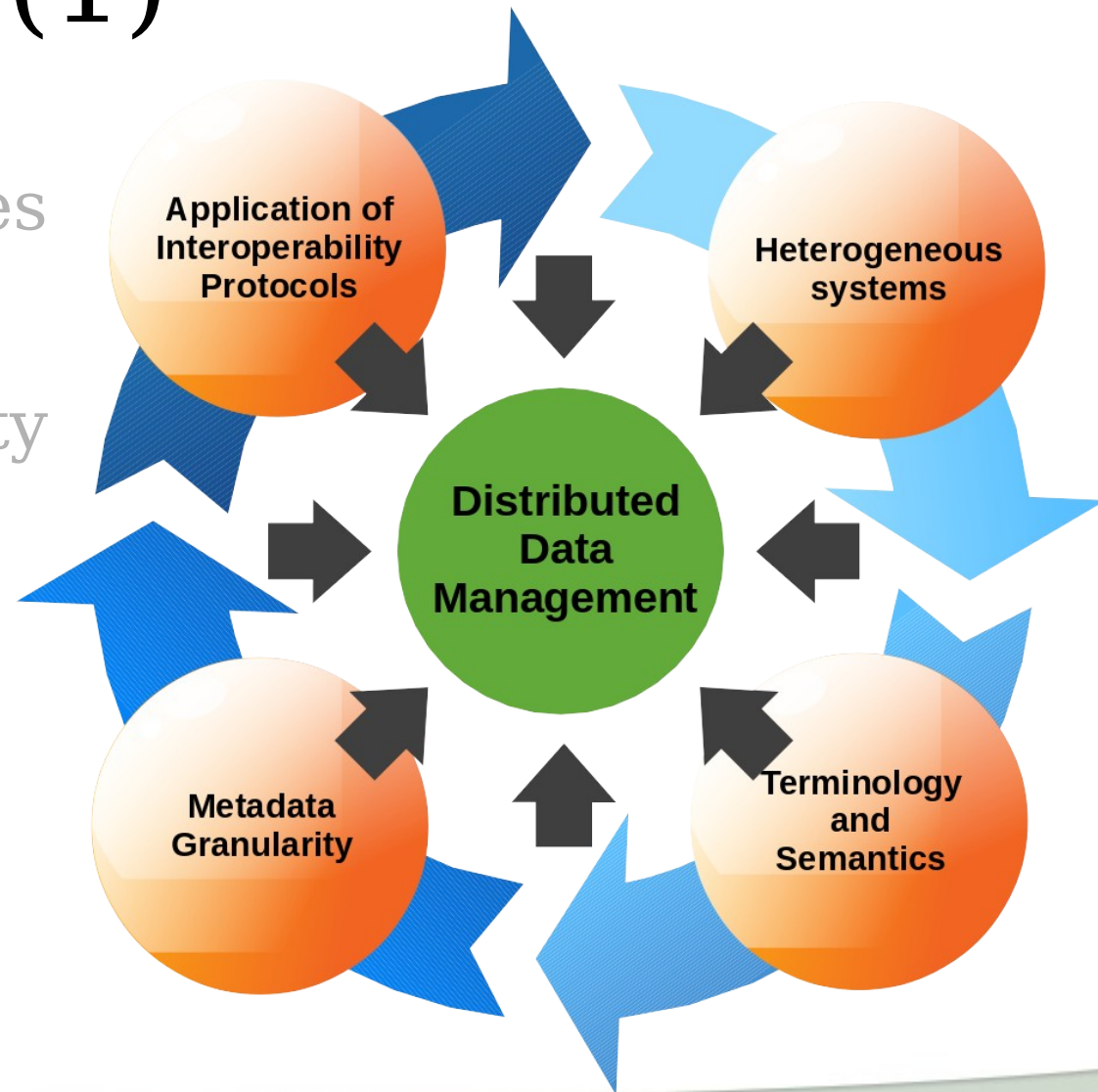
Distributed data management (1)

- Connects physically distributed data centres
- A “virtual” data centre
- Require interoperability
 - Metadata
 - Data
- Powered through documentation of data
 - Discovery metadata
 - Use metadata



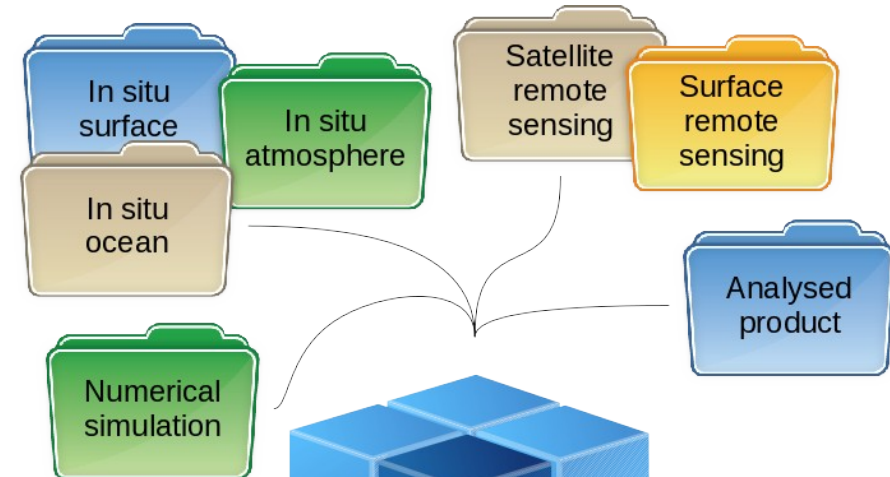
Distributed data management (1)

- Connects physically distributed data centres
- A “virtual” data centre
- Require interoperability
 - Metadata
 - Data
- Powered through documentation of data
 - Discovery metadata
 - Use metadata

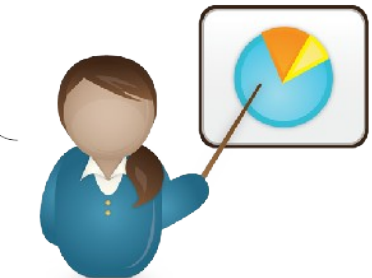


Distributed data management (2)

- Metadata enables qualified access to data and filtering
- Metadata enables independence from applications
- Metadata simplifies integration of different data (data fusion)
- But require a common data model to be cost effective



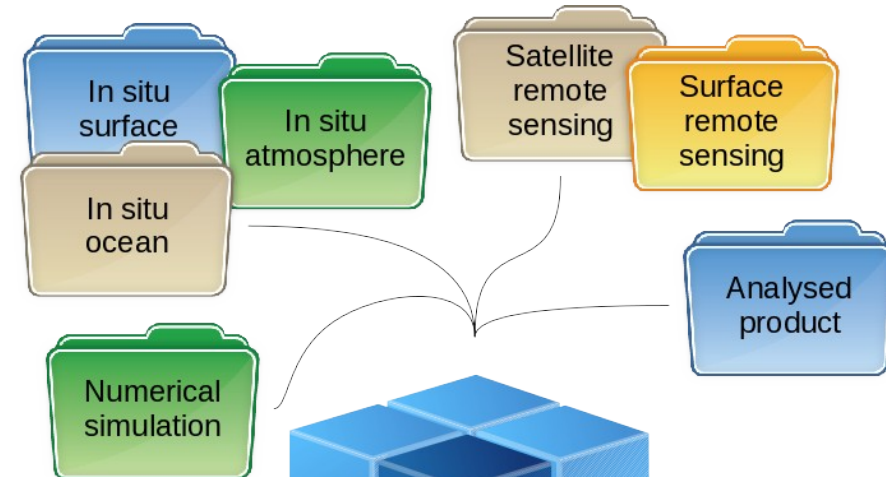
A data model organizes **data** elements and standardizes how the data elements relate to one another.



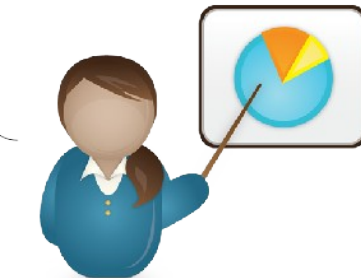
Decision support

Distributed data management (2)

- Metadata enables qualified access to data and filtering
- Metadata enables independence from applications
- Metadata simplifies integration of different data (data fusion)
- But require a common data model to be cost effective



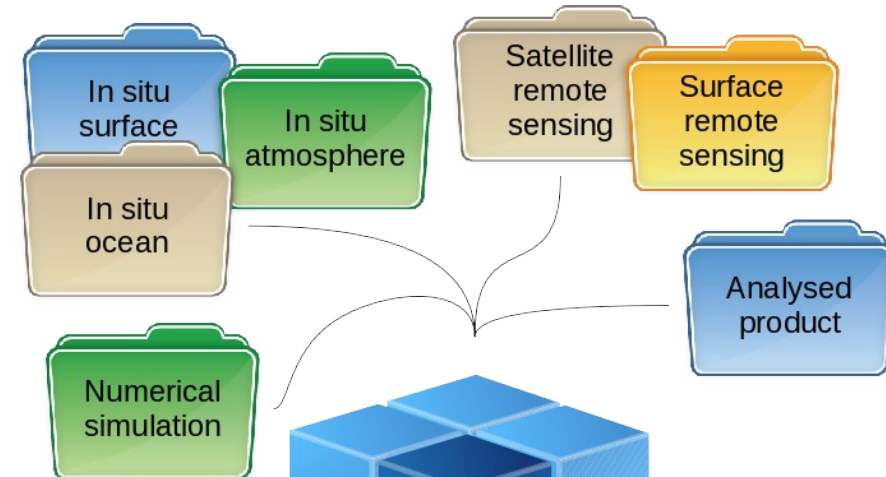
A data model organizes **data** elements and standardizes how the data elements relate to one another.



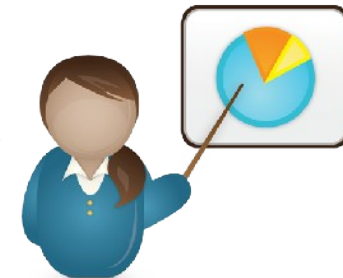
Decision support

Distributed data management (2)

- Metadata enables qualified access to data and filtering
- Metadata enables independence from applications
- Metadata simplifies integration of different data (data fusion)
- But require a common data model to be cost effective



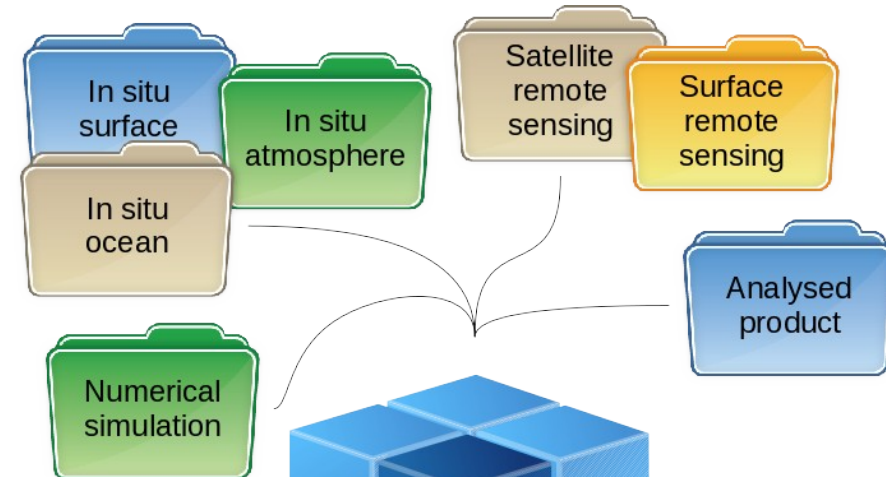
A data model organizes **data** elements and standardizes how the data elements relate to one another.



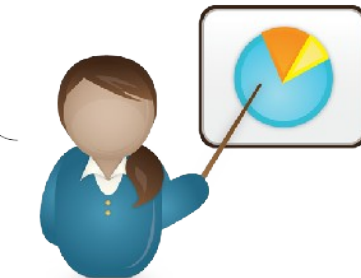
Decision support

Distributed data management (2)

- Metadata enables qualified access to data and filtering
- Metadata enables independence from applications
- Metadata simplifies integration of different data (data fusion)
- But require a common data model to be cost effective



A data model organizes **data** elements and standardizes how the data elements relate to one another.



Decision support

Data access

- All datasets must be available online, provided that no access or use constraints apply.
- User friendly, robust and well known interfaces like HTTP, FTP, OpeNDAP etc will be employed.
- Each dataset should have a unique identifier that can be used to reference the dataset in publications to ensure that data providers receive proper credit for their work, and to allow data citation in scientific journals, in accordance with the norms and guidelines of FORCE11 and DataCite.
- A fourth, basic element is the use of persistent locators (URIs), which constitute a requirement for long term data accessibility.



Data access

- All datasets must be available online, provided that no access or use constraints apply.
- User friendly, robust and well known interfaces like HTTP, FTP, OpeNDAP etc will be employed.
- Each dataset should have a unique identifier that can be used to reference the dataset in publications to ensure that data providers receive proper credit for their work, and to allow data citation in scientific journals, in accordance with the norms and guidelines of FORCE11 and DataCite.
- A fourth, basic element is the use of persistent locators (URIs), which constitute a requirement for long term data accessibility.



Data access

- All datasets must be available online, provided that no access or use constraints apply.
- User friendly, robust and well known interfaces like HTTP, FTP, OpeNDAP etc will be employed.
- Each dataset should have a unique identifier that can be used to reference the dataset in publications to ensure that data providers receive proper credit for their work, and to allow data citation in scientific journals, in accordance with the norms and guidelines of FORCE11 and DataCite.
- A fourth, basic element is the use of persistent locators (URIs), which constitute a requirement for long term data accessibility.



Data access

- All datasets must be available online, provided that no access or use constraints apply.
- User friendly, robust and well known interfaces like HTTP, FTP, OpeNDAP etc will be employed.
- Each dataset should have a unique identifier that can be used to reference the dataset in publications to ensure that data providers receive proper credit for their work, and to allow data citation in scientific journals, in accordance with the norms and guidelines of FORCE11 and DataCite.
- A fourth, basic element is the use of persistent locators (URIs), which constitute a requirement for long term data accessibility.



Plan for access and use, data and knowledge management (1)

- The infrastructure will support scientific projects with data management and long term archival of observations, based on a standard procedure where data providers document, format and upload their datasets according to specifications outlined by NorDataNet.
- The connected repositories will contain both freely accessible data and datasets with access constraints, and will implement procedures to manage the datasets in full accordance with constraints and licenses specified in the metadata.



Plan for access and use, data and knowledge management (1)

- The infrastructure will support scientific projects with data management and long term archival of observations, based on a standard procedure where data providers document, format and upload their datasets according to specifications outlined by NorDataNet.
- The connected repositories will contain both freely accessible data and datasets with access constraints, and will implement procedures to manage the datasets in full accordance with constraints and licenses specified in the metadata.



Plan for access and use, data and knowledge management

(2)

- Data licensing is being developed internationally as a remedy, including tools to attach attribution and licenses physically to the actual data.
- Several such initiatives are relevant to NorDataNet, including those of PIC, Creative Commons, CODATA and NLOD as well as the ongoing work by the Research Council of Norway.
- However, they all pose certain legal problems that will have to be investigated before one or more standard licenses can be applied under the Norwegian jurisdiction.
- To develop an appropriate way of managing data licenses will be a short-term aim of this project.



Plan for access and use, data and knowledge management (2)

- Data licensing is being developed internationally as a remedy, including tools to attach attribution and licenses physically to the actual data.
- Several such initiatives are relevant to NorDataNet, including those of PIC, Creative Commons, CODATA and NLOD as well as the ongoing work by the Research Council of Norway.
- However, they all pose certain legal problems that will have to be investigated before one or more standard licenses can be applied under the Norwegian jurisdiction.
- To develop an appropriate way of managing data licenses will be a short-term aim of this project.



Plan for access and use, data and knowledge management (2)

- Data licensing is being developed internationally as a remedy, including tools to attach attribution and licenses physically to the actual data.
- Several such initiatives are relevant to NorDataNet, including those of PIC, Creative Commons, CODATA and NLOD as well as the ongoing work by the Research Council of Norway.
- However, they all pose certain legal problems that will have to be investigated before one or more standard licenses can be applied under the Norwegian jurisdiction.
- To develop an appropriate way of managing data licenses will be a short-term aim of this project.



Plan for access and use, data and knowledge management (2)

- Data licensing is being developed internationally as a remedy, including tools to attach attribution and licenses physically to the actual data.
- Several such initiatives are relevant to NorDataNet, including those of PIC, Creative Commons, CODATA and NLOD as well as the ongoing work by the Research Council of Norway.
- However, they all pose certain legal problems that will have to be investigated before one or more standard licenses can be applied under the Norwegian jurisdiction.
- To develop an appropriate way of managing data licenses will be a short-term aim of this project.



Linkages

- NorDataNet will be to linked national, European and international data management frameworks and services
 - E.g. GCMD DIF, WIS, NORMAP, NMDC, SIOS KC, GeoAccessNO, ...
- NorDataNet will utilise the national e-infrastructure NorStore as a storage layer for “homeless data”
 - Benefiting from the linkage between storage and processing capabilities through NorStore and NOTUR



Linkages

- NorDataNet will be to linked national, European and international data management frameworks and services
 - E.g. GCMD DIF, WIS, NORMAP, NMDC, SIOS KC, GeoAccessNO, ...
- NorDataNet will utilise the national e-infrastructure NorStore as a storage layer for “homeless data”
 - Benefiting from the linkage between storage and processing capabilities through NorStore and NOTUR