

International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

Y.3601

(05/2018)

SERIES Y: GLOBAL INFORMATION
INFRASTRUCTURE, INTERNET PROTOCOL ASPECTS,
NEXT-GENERATION NETWORKS, INTERNET OF
THINGS AND SMART CITIES

Cloud Computing

**Big data – Framework and requirements for data
exchange**

Recommendation ITU-T Y.3601

ITU-T



ITU-T Y-SERIES RECOMMENDATIONS

GLOBAL INFORMATION INFRASTRUCTURE, INTERNET PROTOCOL ASPECTS, NEXT-GENERATION NETWORKS, INTERNET OF THINGS AND SMART CITIES

GLOBAL INFORMATION INFRASTRUCTURE

General	Y.100–Y.199
Services, applications and middleware	Y.200–Y.299
Network aspects	Y.300–Y.399
Interfaces and protocols	Y.400–Y.499
Numbering, addressing and naming	Y.500–Y.599
Operation, administration and maintenance	Y.600–Y.699
Security	Y.700–Y.799
Performances	Y.800–Y.899

INTERNET PROTOCOL ASPECTS

General	Y.1000–Y.1099
Services and applications	Y.1100–Y.1199
Architecture, access, network capabilities and resource management	Y.1200–Y.1299
Transport	Y.1300–Y.1399
Interworking	Y.1400–Y.1499
Quality of service and network performance	Y.1500–Y.1599
Signalling	Y.1600–Y.1699
Operation, administration and maintenance	Y.1700–Y.1799
Charging	Y.1800–Y.1899
IPTV over NGN	Y.1900–Y.1999

NEXT GENERATION NETWORKS

Frameworks and functional architecture models	Y.2000–Y.2099
Quality of Service and performance	Y.2100–Y.2199
Service aspects: Service capabilities and service architecture	Y.2200–Y.2249
Service aspects: Interoperability of services and networks in NGN	Y.2250–Y.2299
Enhancements to NGN	Y.2300–Y.2399
Network management	Y.2400–Y.2499
Network control architectures and protocols	Y.2500–Y.2599
Packet-based Networks	Y.2600–Y.2699
Security	Y.2700–Y.2799
Generalized mobility	Y.2800–Y.2899
Carrier grade open environment	Y.2900–Y.2999

FUTURE NETWORKS

	Y.3000–Y.3499
--	---------------

CLOUD COMPUTING

Y.3500–Y.3999

INTERNET OF THINGS AND SMART CITIES AND COMMUNITIES

General	Y.4000–Y.4049
Definitions and terminologies	Y.4050–Y.4099
Requirements and use cases	Y.4100–Y.4249
Infrastructure, connectivity and networks	Y.4250–Y.4399
Frameworks, architectures and protocols	Y.4400–Y.4549
Services, applications, computation and data processing	Y.4550–Y.4699
Management, control and performance	Y.4700–Y.4799
Identification and security	Y.4800–Y.4899
Evaluation and assessment	Y.4900–Y.4999

For further details, please refer to the list of ITU-T Recommendations.

Recommendation ITU-T Y.3601

Big data – Framework and requirements for data exchange

Summary

Recommendation ITU-T Y.3601 provides a framework for data exchange in a big data ecosystem. Big data exchange covers multiple processes for data import and data export within a big data ecosystem. Big data exchange is used for exchanging data of multiple types and multiple formats from a data source to a data target. In this Recommendation, direct and intermediary exchange patterns are introduced. In addition, this Recommendation provides a description of the big data activities for the support of big data exchange by extending the activities defined in Recommendation ITU-T Y.3600. Finally, this Recommendation identifies functional requirements that are derived from relevant use cases.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T Y.3601	2018-05-07	13	11.1002/1000/13469

Keywords

Big data, data exchange, framework, requirement, use case.

* To access the Recommendation, type the URL <http://handle.itu.int/> in the address field of your web browser, followed by the Recommendation's unique ID. For example, <http://handle.itu.int/11.1002/1000/11830-en>.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at <http://www.itu.int/ITU-T/ipr/>.

© ITU 2018

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without the prior written permission of ITU.

Table of Contents

	Page
1 Scope.....	1
2 References.....	1
3 Definitions	1
3.1 Terms defined elsewhere	1
3.2 Terms defined in this Recommendation.....	1
4 Abbreviations and acronyms	2
5 Conventions	2
6 Overview of big data exchange	2
6.1 Challenges and benefits of big data exchange.....	3
6.2 General concepts of big data exchange	3
7 Framework of big data exchange.....	5
7.1 Patterns of big data exchange	5
7.2 Extension of activities of big data roles for big data exchange.....	9
8 Functional requirements of big data exchange	11
8.1 Requirements for data registration and cataloguing.....	11
8.2 Requirements for data retrieval	12
8.3 Requirements for data delivery	13
8.4 Requirements for customer support.....	13
8.5 Requirements for data quality management	13
8.6 Requirements for data rights management	14
8.7 Requirements for management of personal information.....	14
9 Security considerations	14
Appendix I – General procedure for data exchange	15
Appendix II – Use cases of big data exchange	16
II.1 Big data direct exchange	16
II.2 Big data intermediary exchange	17
Bibliography.....	31

Recommendation ITU-T Y.3601

Big data – Framework and requirements for data exchange

1 Scope

This Recommendation specifies the framework and requirements for data exchange in a big data ecosystem. This Recommendation identifies general concepts, patterns, activities, and functional requirements based on the big data ecosystem and capabilities defined in [ITU-T Y.3600]. The functional requirements provided in this Recommendation are derived from use cases.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

[ITU-T Y.3600] Recommendation ITU-T Y.3600 (2015), *Big data – Cloud computing based requirements and capabilities*.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 activity [b-ITU Y.3502]: A specified pursuit or set of tasks.

3.1.2 big data [ITU-T Y.3600]: A paradigm for enabling the collection, storage, management, analysis and visualization, potentially under real-time constraints, of extensive datasets with heterogeneous characteristics.

NOTE – Examples of datasets characteristics include high-volume, high-velocity, high-variety, etc.

3.1.3 metadata [b-ISO/IEC 2382]: Data about data or data elements, possibly including their data descriptions, and data about data ownership, access paths, access rights and data volatility.

3.1.4 role [b-ITU Y.3502]: A set of activities that serves a common purpose.

3.1.5 sub-role [b-ITU Y.3502]: A subset of the activities of a given role.

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 big data exchange: Multiple processes consisting of data import and data export within a big data ecosystem.

NOTE – A big data ecosystem defines necessary activities for roles providing and consuming big data services as well as relationships between roles (see [ITU-T Y.3600]).

3.2.2 data catalogue: A listing of all metadata which a data broker makes available.

3.2.3 data export: A process for delivering data.

3.2.4 data import: A process for receiving data.

3.2.5 entity: A grouping of big data roles and sub-roles that can be played by a party.

3.2.6 processed data: Data produced by processing steps in a big data ecosystem.

NOTE – Processing steps include data collection, data preparation, data visualisation and data analysis.

3.2.7 raw data: Data from a data source without any alteration.

NOTE – Raw data is also known as unprocessed data.

3.2.8 service catalogue: A listing of all the big data services of a particular big data service provider.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

API Application Programming Interface

BDC Big Data service Customer

BDSP Big Data Service Provider

DP Data Provider

5 Conventions

In this Recommendation:

The keywords "**is required**" indicate a requirement which must be strictly followed and from which no deviation is permitted if conformance to this document is to be claimed.

The keywords "**is recommended**" indicate a requirement which is recommended but which is not absolutely required. Thus this requirement need not be present to claim conformance.

The keywords "**can optionally**" indicate an optional requirement which is permissible, without implying any sense of being recommended. This term is not intended to imply that the vendor's implementation must provide the option and the feature can be optionally enabled by the network operator/service provider. Rather, it means the vendor may optionally provide the feature and still claim conformance with the specification.

In the body of this document, the words shall, shall not, should, and may sometimes appear, in which case they are to be interpreted, respectively, as is required to, is prohibited from, is recommended, and can optionally. The appearance of such phrases or keywords in an appendix or in material explicitly marked as informative are to be interpreted as having no normative intent.

6 Overview of big data exchange

Data exchange is a process of:

- receiving source data under a source schema from a data source;
- transforming the received source data into target data under a target schema without altering the representation of the source data; and
- delivering the target data to the data target.

Appendix I provides a high-level view regarding data exchange procedures between two systems.

When applied to a big data ecosystem, data exchange will typically involve exchange of data among different data sources (e.g., data providers) and data targets (e.g., big data service customers).

In the following, the term "big data exchange" is used to refer to "data exchange" in a big data ecosystem. Big data exchange involves multiple processes including data import and data export. Big

data exchange enables exchanging data of multiple types and multiple formats from a data source to a data target. Characteristics of exchanged data are as follows:

- types of exchanged data include structured data, semi-structured data and unstructured data;
- formats of exchanged data include text, spreadsheet, video, audio, image, geographical position, map, and combinations of afore-mentioned formats (such as web documents, sensing data, media streaming);
- exchanged data are categorized as raw (i.e., unprocessed) data and processed data.

6.1 Challenges and benefits of big data exchange

The following challenges are important to be considered for big data exchange:

- **various sources, types and formats of data:** big data service providers have to handle diverse aspects of data and data sources during data collection, storage and integration;
- **schema-on-read:** usually big data is stored in a raw format, but after data is discovered and captured, it is transformed to fulfil application's requirements;
- **unawareness of suitable data/Unconstrained usage of data:** sometimes, the big data service customer does not recognize what kinds of data are really needed caused by unconstrained usage of data in big data ecosystem. Therefore, data provider or big data service provider should offer data with their usage for big data service customer to choose the helpful data for resolving their problem.

Exchanging data in a big data ecosystem is expected to provide the following benefits:

- **mitigation of silos** in the ecosystem through a better sharing of high-variety data between involved parties;
- **monetization of data** enabling better revenues to be made by parties from high volume of data exchanged in the ecosystem;
- **openness of the publicly available data** contributing to human society and economic activities;
- **facilitation of the appearance** about new and effective business models;
- **interconnection of valuable, high-variety, and high-volume data** contributing more to human society and economic activities.

6.2 General concepts of big data exchange

Figure 6-1 illustrates the primitive model for a big data exchange. In this model, the following principles apply:

- a data source and a data target communicate with each other through a "data exchange". Through this (illustrated by the black arrow in Figure 6-1), data is exchanged from the data source to the data target. During this exchange of data, data processing may be performed;
- the data source is an entity which collects datasets (including raw data or already processed data) or big data service's output and exports them to the data target;
- in the relationship, the data target is an entity which imports dataset (including raw data or processed data) or big data service's output from the data source. Data export from the data source is triggered either by the big data source itself or by an initiation request received from the data target.



Figure 6-1 – Primitive model of big data exchange

Big data exchange patterns result from a composition of the primitive model Figure 6-2 illustrates two major patterns:

- **direct exchange pattern:** a direct exchange of data from a peer data source to a peer data target (see clause 7.1.1);
- **intermediary exchange pattern:** an indirect exchange of data through an intermediary control and data processing agent (e.g., an intermediary) (see clause 7.1.2).

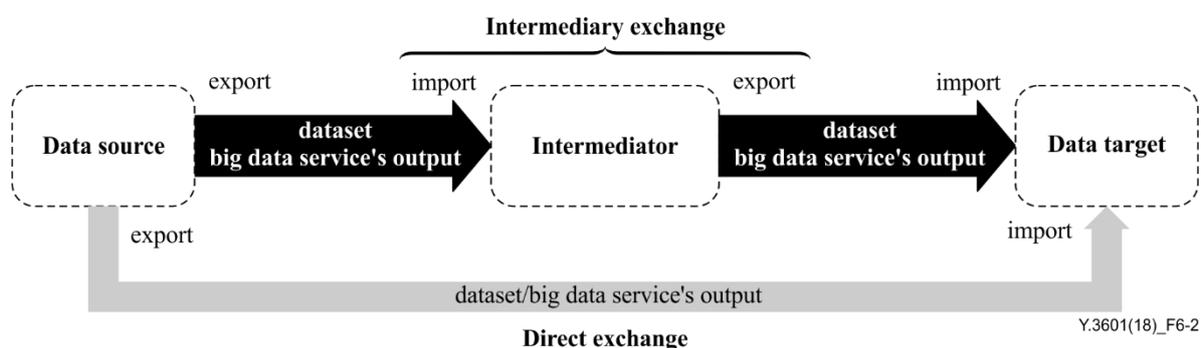


Figure 6-2 – Major patterns of big data exchange

As shown in Figure 6-2, the major patterns of big data exchange rely on the use of three entities:

- 1) **data source:** An entity which collects dataset and big data service's output and exports them to the intermediary (in the case of intermediary exchange) or to the data target (in the case of direct exchange);
- 2) **intermediator:** An entity which imports datasets or big data service's output from a data source and exports them as datasets or big data service's output to a data target;

NOTE 1 – The intermediary may process raw datasets and produce the processed dataset by data processing, such as data collection, preparation, analysis and visualization.

- 3) **data target:** An entity which requests the intermediary (in case of intermediary exchange) or the data source (in case of direct exchange) for a dataset or a big data service's output and imports a dataset or big data service's output from the intermediary or the data source.

Examples of big data exchange relationships based on general concept aligning in a big data ecosystem are shown in Figure 6-3.

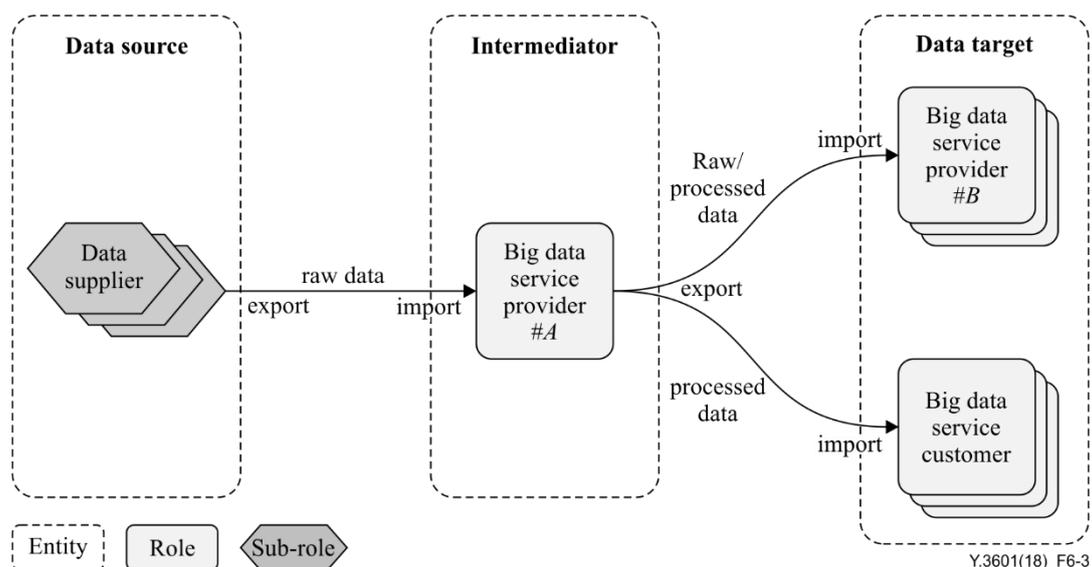


Figure 6-3 – Example of big data exchanges in a big data ecosystem

As shown in Figure 6-3:

- data suppliers export raw data to big data service provider #A; Each data supplier acts as a data source while big data service provider #A acts as an intermediary;
- big data service provider #A imports raw data from data suppliers, processes these raw data and outputs processed data. Big data service provider #A exports either raw data and/or processed data to big data service provider #B and/or exports processed data to big data service customer;
- big data service provider #B imports raw data and/or processed data from big data service provider #A. Big data service provider #B makes use of these imported data;
- big data service customer imports processed data from big data service provider #A and makes use of these imported data.

NOTE 2 – Although not shown in the figure, data supplier is a sub-role of data provider as per [ITU-T Y.3600].

NOTE 3 – Big data service providers in big data exchange expose the availability of processing capabilities and export the captured, collected, prepared, analysed and visualized data to big data service customers on demand.

7 Framework of big data exchange

This clause describes the framework of big data exchange including further details on the two major big data exchange patterns introduced in clause 6 and the description of additional activities to the ones defined [ITU-T Y.3600].

7.1 Patterns of big data exchange

7.1.1 Direct exchange pattern

A direct exchange pattern is represented by a primitive model (see clause 6.2), which means it is composed of interactions between a data source and a data target. In the big data ecosystem, there are three types of interaction scenarios available for the direct exchange pattern as shown in Figure 7-1.

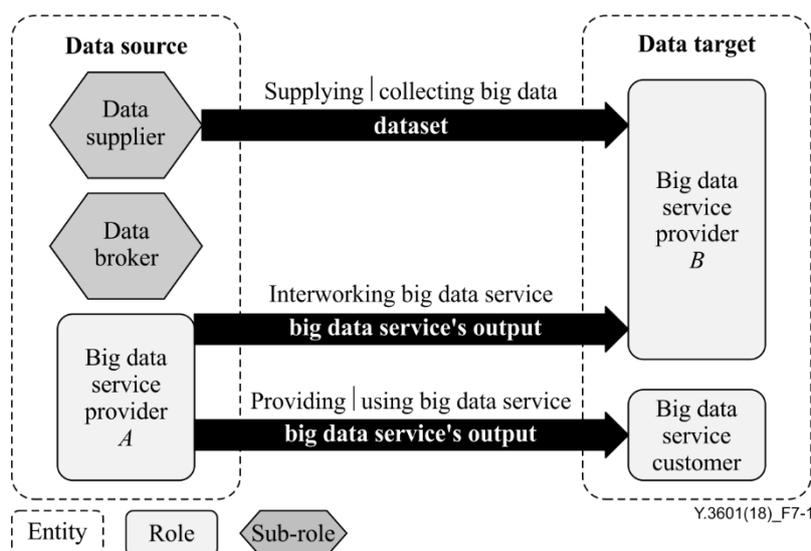


Figure 7-1 – Direct exchange

The three interaction scenarios of direct exchange pattern are as follow:

- 1) **EXsc**: Supplying | collecting big data (Data supplier → Big data service provider): a data supplier provides datasets to a big data service provider by using on-line or off-line methods, and big data service provider uses them;
- 2) **EXiw**: Interworking big data service (Big data service provider → Big data service provider): a big data service provider uses big data service's output from a peer big data service provider;
- 3) **EXpu**: Providing | using big data service (Big data service provider → Big data service customer): a big data service customer uses big data service's output from a big data service provider.

NOTE – Although not shown in the figure, data supplier and data broker are a sub-role of data provider as per [ITU-T Y.3600].

In Figure 7-1, the data broker provides meta-information (data catalogue for dataset) registry to data source (Data supplier and Big data service provider A), and also provides a service catalogue to the data target (Big data service provider B and Big data service customer). The data target searches data from the data broker on the data source's side.

A data source publishes metadata to an online registry for data sharing. In Figure 7-2, the data supplier and the big data service provider (Big data service provider A) on the data source's side publish metadata and big data service catalogue to the external data broker. The big data service provider (Big data service provider B) and the Big data service customer on the data target's side search data or big data service on the external data broker and use them.

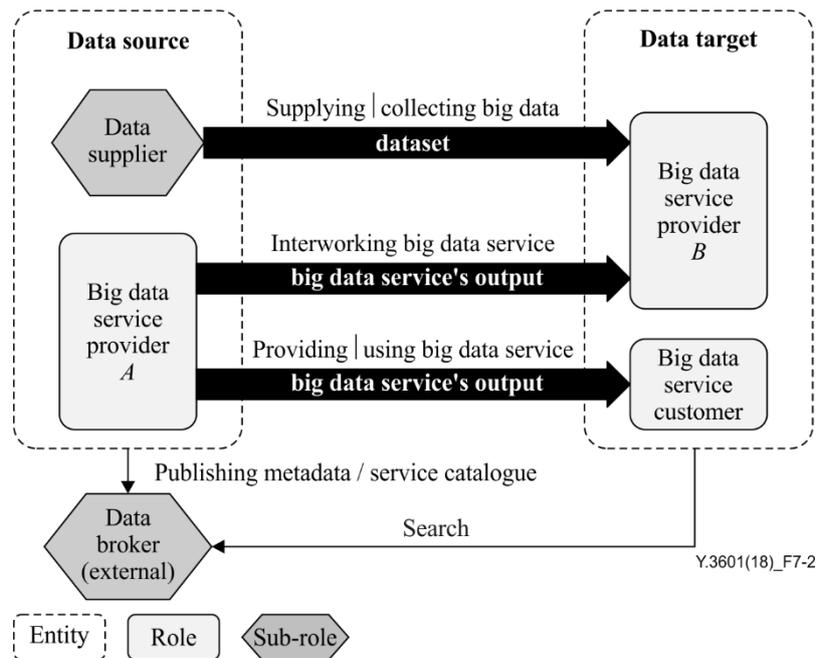


Figure 7-2 – Direct exchange using an external data broker

7.1.2 Intermediary exchange pattern

The intermediary exchange pattern is composed of a set of direct exchange patterns. In Figure 7-3, for each direct exchange pattern involving a data source, the intermediary entity is acting as data target while for each direct exchange pattern involving a data target, the intermediary entity is acting as a data source.

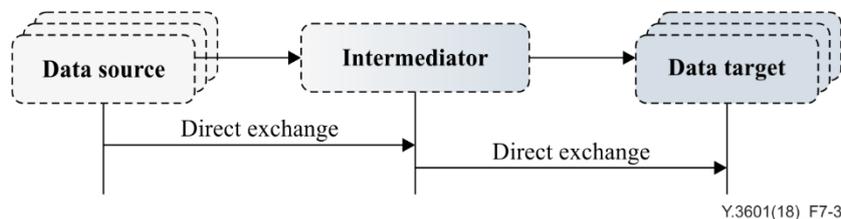


Figure 7-3 – Intermediary exchange

In an intermediary exchange, the intermediary may modify data received from data sources by pre-processing, integrating, analysing and visualising data.

NOTE – In Figure 7-4, the intermediary entity is acting as a big data service provider and a data broker.

Figure 7-4 illustrates the case where the intermediary imports and manipulates data from multiple data sources (Data supplier and Big data service provider *S*). Data broker *I* in the intermediary provides meta-information registry. The big data service provider *I* provides a metadata and service catalogue to the data target.

Between a data source and an intermediary, two types of exchange interaction scenarios are shown as follows:

- 1) **EX_{sc}**: Data supplier → Big data service provider;
- 2) **EX_{iw}**: Big data service provider → Big data service provider.

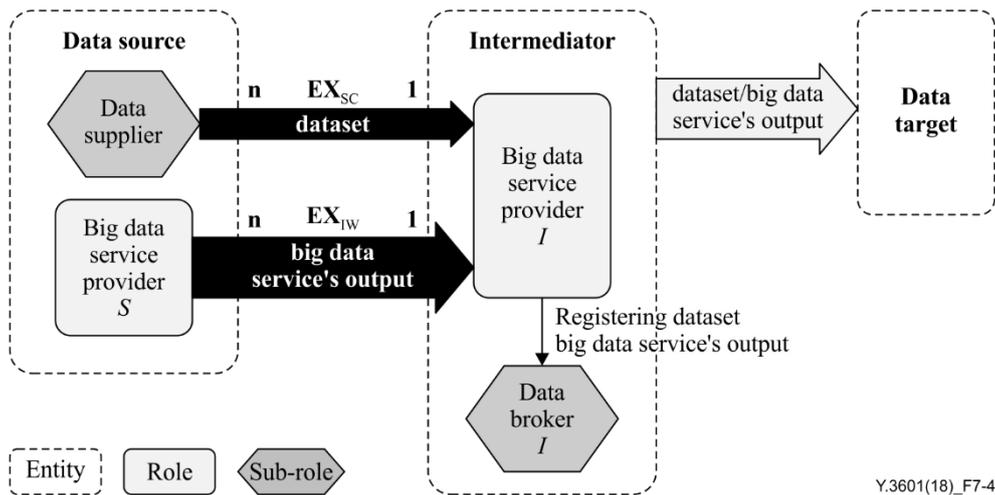


Figure 7-4 – Import of intermediary exchange pattern

Between the intermediary and the data target in Figure 7-5, data broker *I* in the intermediary provides a metadata and service catalogue to the data target, and three types of scenarios (EX_{SC}, EX_{IW}, EX_{PU} in clause 7.1.1) are used.

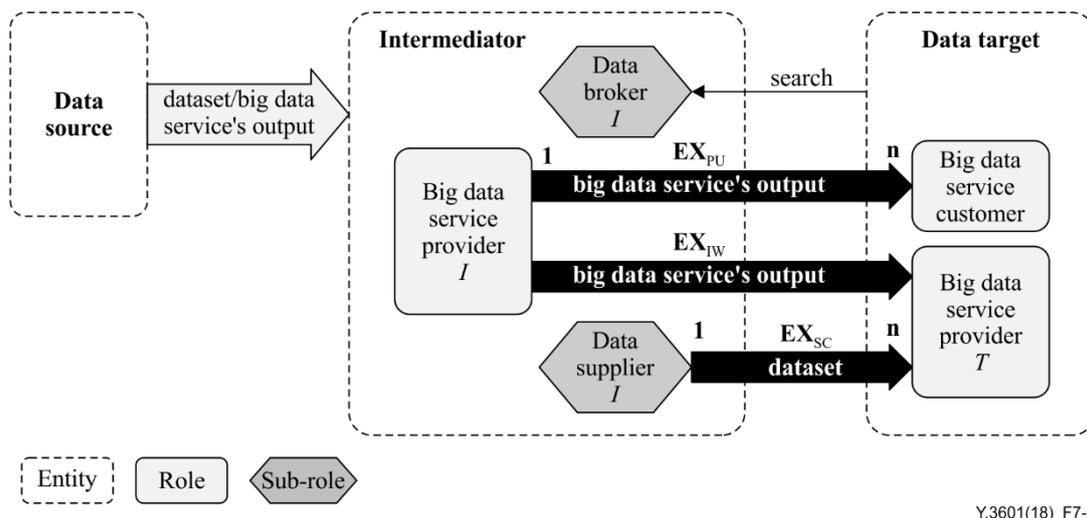
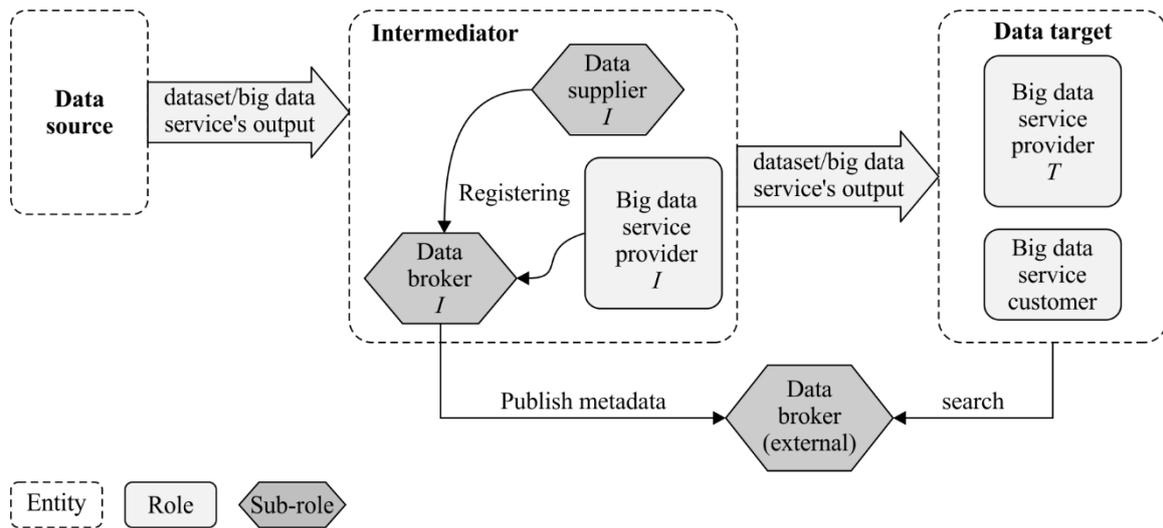


Figure 7-5 – Export of intermediary exchange pattern using internal data broker

The case of intermediary exchange pattern, intermediary may publish meta-information to the external online registry for data sharing. In Figure 7-6, Data supplier *I* and Big data service provider *I* publish metadata and big data service catalogue to data broker *I*. Data broker *I* re-publishes the information to the external data broker, and the data target searches data or big data service on the external data broker and requests them to the intermediary.

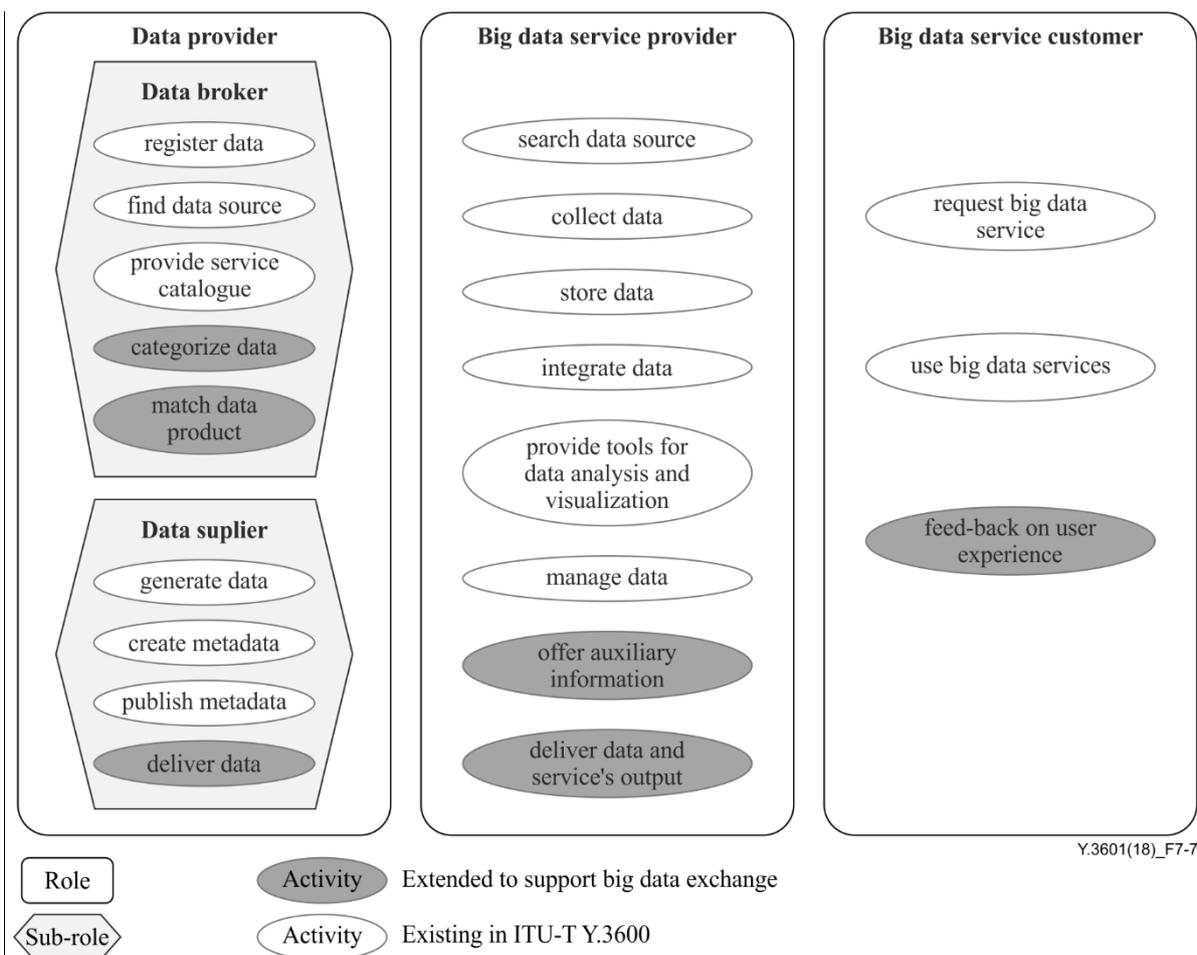


Y.3601(18)_F7-6

Figure 7-6 – Export of intermediary exchange pattern using external data broker

7.2 Extension of activities of big data roles for big data exchange

Big data exchange inherits the big data ecosystem defined in [ITU-T Y.3600]. Several activities are extended to support big data exchange.



Y.3601(18)_F7-7

Figure 7-7 – Extended activities of big data for big data exchange

NOTE – Figure 7-7 identifies reused activities from [ITU-T Y.3600] as well as new activities needed to support big data exchange.

7.2.1 Data provider

The data provider (DP) roles consist of two-sub-roles, data supplier and data broker (see clause 6.2.1 in [ITU-T Y.3600]).

7.2.1.1 Data supplier

In case of direct exchange, data supplier sub-role is performed by the data source entity while in intermediary exchange data supplier sub-role can be performed by the intermediary.

The existing activities of data supplier in [ITU-T Y.3600] for big data exchange are provided:

- generate data;
- create metadata (information);
- publish metadata (information).

The additional activities of data supplier for big data exchange include:

- **deliver data:** transfer data.

NOTE – Delivery methods depend on data transfer context. DP:data supplier can send data in real-time or by batch updates. Also, DP:data supplier can use synchronous data transfer rather than asynchronous data transfer for time sensitive service.

7.2.1.2 Data broker

In case of a direct exchange pattern, the data broker sub-role is performed by the data source entity while in case of an intermediary exchange pattern, the data broker sub-role is performed by the intermediary. Also, data broker can be located outside of the data source and data target.

Activities of data broker defined in [ITU-T Y.3600] are reused as follows for big data exchange:

- **register data;**

NOTE 1 – This activity is renamed from "providing a meta-information registry to the data supplier for publishing their data sources" in [ITU-T Y.3600].

- **find data source;**

NOTE 2 – This activity is renamed from "finding on-line open-data sources and registering corresponding meta-information" in [ITU-T Y.3600].

- **provide service catalogue.**

NOTE 3 – This activity is renamed from "providing a service catalogue to the big data service provider for searching usable data" in [ITU-T Y.3600].

NOTE 4 – A service catalogue includes "data catalogue".

New activities as compared to data broker activities in [ITU-T Y.3600] include:

- **categorize data:** process of organizing data into categories for its effective and efficient use;

NOTE 5 – It supports data classification either by common terminology and taxonomy, as well as application fields.

- **match data:** match and recommend data to a big data service provider based on multiple criteria.

7.2.2 Big data service provider

In case of direct exchange, big data service provider is performed by the data source and the data target entities. While in case of an intermediary pattern, the big data service provider is performed by the data source, the intermediary, and the data target.

Activities of big data service provider defined in [ITU-T Y.3600] are reused with modifications for big data exchange as follows:

- **search data source;**

NOTE 1 – This activity is split from "searching data sources and collecting data by requesting and crawling" in [ITU-T Y.3600].

- **collect data;**

NOTE 2 – This activity is split from "searching data sources and collecting data by requesting and crawling" in [ITU-T Y.3600].

- **store data;**

NOTE 3 – This activity is renamed from "store data to data repository" in [ITU-T Y.3600].

- **integrate data;**

NOTE 4 – This activity includes mash-up data on demand.

- **provide tools for data analysis and visualization;**

NOTE 5 – This activity includes configuration of big data service from a BDC side; for example, modification of big data analyzing algorithms.

- **manage data.**

NOTE 6 – This activity includes managing data quality, data provenance, protection of personal information, security, policy, ownership, licensing, etc.

New activities of big data service provider as compared to [ITU-T Y.3600] include:

- **offer auxiliary information:** provide auxiliary information about a dataset or a big data service to improve service utilization;

NOTE 7 – Auxiliary information consists of statistics on summary of usage, data quality or data provenance, etc.

- **deliver data and service's output:** transfer either datasets or big data service's output.

NOTE 8 – Data delivery methods depend on service context. BDSP can send data in real-time or by batch updates. Also, BDSP can use synchronous data transfer rather than asynchronous data transfer for time sensitive service.

7.2.3 Big data service customer

In the concept of big data exchange, the big data service customer is performed by the data target entity.

Activities of big data service customer defined in [ITU-T Y.3600] are reused for big data exchange as follows:

- **request big data service;**

NOTE 1 – This activity is renamed from "requesting big data services from the big data service provider" in [ITU-T Y.3600].

- **use big data service;**

NOTE 2 – This activity is renamed from "use the output of big data services" in [ITU-T Y.3600].

New activities of big data service customer as compared to [ITU-T Y.3600] include:

- **provide feed-back on user experience:** evaluate user experience on big data service.

NOTE 3 – User experience includes service satisfaction or complaint, usage of service, etc.

8 Functional requirements of big data exchange

This clause describes the functional requirements of big data exchange.

8.1 Requirements for data registration and cataloguing

Data registration and cataloguing requirements include:

- 1) it is required that DP:data broker provides a common data catalogue schema which cover various type of data;
NOTE 1 – A common data catalogue includes data type, format, size, category, metadata information and its URI, delivery mechanism, update frequency, electronic access methods such as API. This catalogue may include data rights, license policy, price, quality of data, aggregation information, pre-processing information, usage of data, related keywords, and sample data.
- 2) it is required that DP:data broker provides data catalogue registration mechanisms to DP:data supplier;
NOTE 2 – Registration mechanisms are user interface or API for registration in forms of open API.
- 3) it is recommended that DP:data supplier supports automatic extraction of the information from data in order to provide the associated metadata to be included in data catalogues;
- 4) it is recommended that DP:data broker provides notifications of newly registered metadata and manages subscriptions to such notifications;
NOTE 3 – To improve data distribution and utilization, DP:data broker notifies the newly registered data to BDSP as a task of publish data.
- 5) it is required that DP:data broker supports data classification with commonly used data vocabulary and taxonomy;
- 6) DP:data broker can optionally support multiple application domain specific vocabularies and taxonomies for a single source of data;
- 7) it is recommended that DP:data broker supports multiple data classifications by its areas of use;
- 8) it is required that DP:data broker supports the publication of data specifications to BDSPs;
NOTE 4 – A data specification is used for manipulating data from storage or streaming data in an on-demand manner (e.g., triggered when a user is requesting data). It includes information about source of data, a process of generating data, selling policies for data, etc.
- 9) it is recommended that DP:data broker performs verification of data before publishing these data.
NOTE 5 – The DP:data supplier registers metadata to the data catalogue of the DP:data broker to make the data available for distribution.
NOTE 6 – For the verification phase, the DP:data broker may request the additional information about the data and sampling data from the DP:data supplier.

8.2 Requirements for data retrieval

Data retrieval requirements include:

- 1) it is required that DP:data broker provides the BDSP with an interface to access data catalogues;
NOTE 1 – This interface may support user specific taxonomy (e.g., Web Ontology Language file) to extend a keyword search.
- 2) it is required that DP:data broker provides metadata searching functionalities to BDSP;
NOTE 2 – Examples of search method are keyword search and directory search.
NOTE 3 – The results of search can be listed by registered date, price, sales ranking, and data supplier's credit according to the metadata schema.
- 3) it is recommended that DP:data broker recommends data to BDSP based on multiple criteria;
NOTE 4 – The criteria include price, discount rate, application category, etc.
- 4) it is recommended that DP:data broker supports the best match searching with a data request from BDSP.

8.3 Requirements for data delivery

Data delivery requirements include:

- 1) it is required that BDSP provides effective data transmission methods;
NOTE 1 – Effective data transmission methods include secure connection control, network consistency & scalability, and weight lightening (e.g., data compression), etc.
- 2) it is required that BDSP requests a retransmission when the data transmission fails;
- 3) it is recommended that BDSP supports various data delivery methods;
NOTE 2 – Data delivery method includes download at once, online cloud storage service, a form of database (e.g., Structured Query Language interface), an event push, etc.
- 4) it is required that DP:data broker mediates the secure transmission mechanism between the data provider and the big data service provider;
NOTE 3 – An example of mediation of secure transmission is to provide public key for data provider and big data service provider.
- 5) it is recommended that BDSP supports the combination, aggregation, integration and intermediation of datasets from multiple data sources;
- 6) it is recommended that BDSP assembles data according to published data specifications;
NOTE 4 – A data specification includes information about real-time data material with data access information, collection information, and information about a form in which data are stored, etc. In this case, BDSP connects to DP:data supplier. The BDSP requests data from DP:data supplier and stores the received data in order to perform data assembling.
- 7) it is required that DP:data supplier provides mechanisms to protect integrity, availability and authenticity of data;
- 8) it is recommended that DP:data broker provides charging methods on the data.

8.4 Requirements for customer support

Customer support requirements include:

- 1) it is required that BDSP provide interfaces to BDC for operating a workflow and data analysis;
NOTE 1 – Workflow is the series of analysing events that are necessary to complete a task. BDC can utilize the workflow for automating processes related to the analysis of big data.
- 2) it is recommended that BDSP supports a method for utilizing a user defined analysis algorithm;
NOTE 2 – A method for utilizing a user defined analysis algorithm includes algorithm development functions, a common model to describe analysis algorithm, and an encoding tool for converting them to a common model, etc.
- 3) it is recommended that BDSP provides a meta-information for an analysed result to BDC.
NOTE 3 – A meta-information may include metadata of analysed data, and functions with parameters which had been applied.

8.5 Requirements for data quality management

Data quality management requirements include:

- 1) it is recommended that BDSP provides an evaluation method for managing data quality;
NOTE – Evaluation methods includes scoring (rating), reviewing, etc.
- 2) it is recommended that BDSP provides evaluation result about data quality;
- 3) it is recommended that DP:data broker is able to access that the quality of data provided by the DP:data supplier is reliable.

8.6 Requirements for data rights management

Data rights management requirements include:

- 1) it is required that DP:data supplier supports the deletion of datasets when the data ownerships is transferred;
- 2) it is required that BDSP supports the deletion of data when the ownerships of a data were transferred to BDC;
- 3) it is recommended that DP:data broker supports registration of the data rights with metadata;
NOTE 1 – Data rights contains information of general agreement, copyrights, patents, licenses policies and pricing policies.
- 4) it is recommended that DP:data broker supports search of data rights;
- 5) it is recommended that BDSP supports the storage for metadata as well as data rights.
NOTE 2 – Data rights contains information of general agreement, copyrights, patents, licenses policies and pricing policies.

8.7 Requirements for management of personal information

Personal information management requirements include:

- 1) it is required that BDSP supports deletion customer's data by the policy when the service is terminated;
NOTE – Customer's data includes source data, result data, and applied analytical methods, etc.
- 2) it is required that BDSP supports data de-identification based on personal information management policy.

9 Security considerations

Relevant security requirements of [b-ITU-T Y.2201], [b-ITU-T Y.2701] and applicable X, Y and M series of ITU-T Recommendations need to be taken into consideration, including access control, authentication, data confidentiality, data retention policy, network security, data integrity, availability and protection of personal information.

Appendix I

General procedure for data exchange

(This appendix does not form an integral part of this Recommendation.)

This appendix provides a general procedure of data exchange.

A data exchange context is the framework in which two or more systems agree to exchange identified sets of data and related attributes.

Concerning the interoperability of data exchange between systems, the following two fundamental issues have to be considered:

- the semantics of the content and the logical structures of data (application schema);
- a system and platform-independent data structure that can represent data corresponding to the application schema (for example a vendor neutral or well-known data format).

NOTE 1 – An application schema is a conceptual schema for applications which defines content and structure of the data (e.g., eXtensible Markup Language (XML) based metadata interchange).

NOTE 2 – Data exchange between homogeneous systems (which are using the same format and/or schema) means fewer interoperability issues to be solved than data exchange between heterogeneous systems (in which the used format and schema are different).

Figure I.1 provides a procedure of data exchange between two systems.

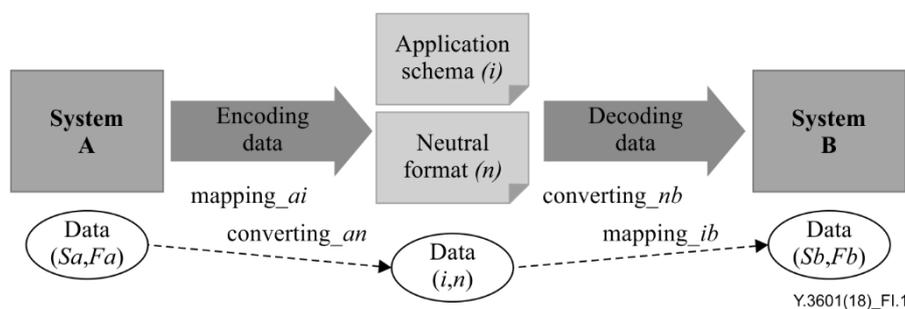


Figure I.1 – Data exchange procedure

Between System A (which has source schema sa , data format Fa) and System B (which has target schema sb , data format Fb), the data exchange takes place with the following two steps:

- encoding data: this includes mapping the Sa source schema onto application schema i (cf. "mapping_ai" in the figure), and converting the data format Fa to n (cf. "converting_an" in the figure);
- decoding data: this includes converting the data format n to Sb (cf. "converting_nb" in the figure), and mapping application schema i with target schema B (cf. "mapping_ib" in the figure).

In general, the following preconditions have to be considered concerning data exchange:

- the application domains of each system are same or similar;
- the application schemas of each system are similar enough for allowing conversion from one to the other.

Appendix II

Use cases of big data exchange

(This appendix does not form an integral part of this Recommendation.)

II.1 Big data direct exchange

II.1.1 Publishing various kinds of data to a data broker

Title	Publishing various kinds of data to a data broker (DP:data broker)
Description	<p>A company would like to sell various kinds of its own data (becoming thus a data supplier). In order to promote selling, a person in charge in the company finds a brokerage service (provided by a data broker) supporting a connection between supply-side and demand-side of data. Then, he uploaded the metadata to on-line brokerage service.</p> <p>When the data supplier uploads (or fills in) the requested form with a short data description or data category described with key words, the data broker categorizes the data by mapping with the supported taxonomies.</p>
Roles/Sub roles	<ul style="list-style-type: none"> – DP:data supplier – DP:data broker
Figure (optional)	<p>The diagram illustrates the process of publishing data to a data broker. At the top, a dashed box labeled 'Data source' is connected by dotted lines to a 'Data supplier' (represented by a hexagon) and a 'Data broker' (represented by a hexagon). A solid arrow labeled 'publish data' points from the Data supplier to the Data broker. A curved arrow labeled 'categorize data' points from the Data broker to a 'Related domain' box. This box contains a list of categories: 'Health', 'Shopping', and 'Hobby', each with a checked checkbox, followed by two empty checkboxes with ellipses. The diagram is captioned 'Y.3601(18)_TableII.1-1'.</p>
Pre-conditions (optional)	<p>The DP:data broker provides a service interface allowing the DP:data supplier to registering the various kinds of data.</p> <p>The DP:data broker provides multiple domain categories with single data source.</p>
Post-conditions (optional)	
Derived requirement	<ul style="list-style-type: none"> – Data registration and cataloguing (refer to clause 8.1 requirements 1, 2, 3, 5, 6)

II.1.2 Online big data market place using a data broker based on the trust

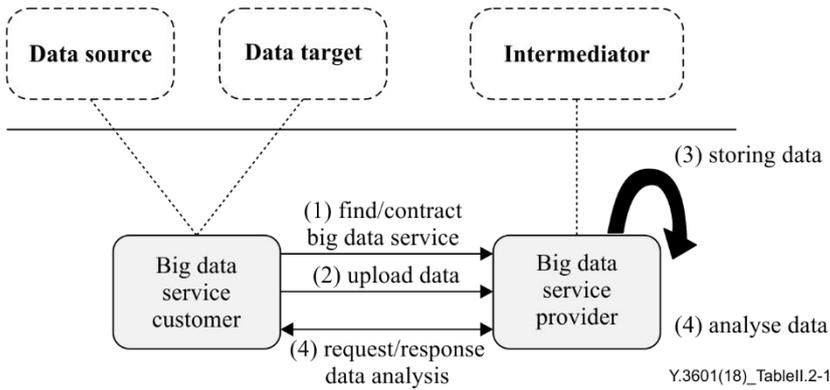
Title	Online big data market place using a data broker based on the trust
Description	<p>In a big data market place, a data catalogue hosted by a data broker is used for providing information about data that can be delivered by data suppliers. Information in the catalogue includes data type, format, size, update frequency, delivery mechanism, category, price, etc. This data catalogue is consulted by big data service providers (BDSPs) to purchase data from data suppliers.</p> <p>The data market place constitutes a trusted environment between data suppliers, the data broker and BDSPs. Verification of the data in terms of trust and secure transfer of data from seller to buyer is expected in the big data market place in order to provide data reliability and secure data transmission.</p>

Title	Online big data market place using a data broker based on the trust
Roles/Sub roles	<ul style="list-style-type: none"> – DP:data supplier (big data seller) – DP:data broker (big data market place) – BDSP (big data buyer)
Figure (optional)	
Pre-conditions (optional)	The data broker mediates between the data suppliers and the big data service providers that are part of the market place. The data broker does not deliver data to BDSPs: this is the role of data suppliers.
Post-conditions (optional)	
Derived requirement	<ul style="list-style-type: none"> – Data registration and cataloguing (refer to clause 8.1 requirements 2, 9) – Data retrieval (refer to clause 8.2 requirement 1) – Data delivery (refer to clause 8.3 requirements 2, 4, 8) – Data quality management (refer to clause 8.5 requirement 3)

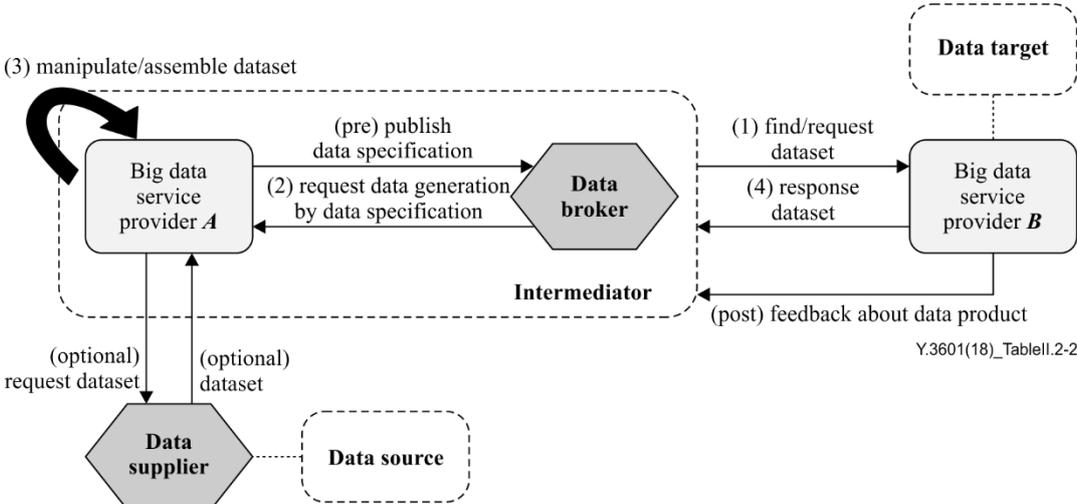
II.2 Big data intermediary exchange

II.2.1 Use a public big data service for analysing customer's own big data

Title	Use a public big data service for analysing the customer's own big data
Description	A company would like to use an online big data service for analysing its own big data (since not able to handle such data by its own means). Therefore, the company searches and finds a suitable online big data service from a BDSP for this purpose. Thanks to the big data service, the company (being the big data service customer (BDC)) uploads its own data, and interacts with the BDSP through a provided user interface to perform an analysis of the uploaded data.
Roles/Sub roles	<ul style="list-style-type: none"> – BDC – BDSP

Title	Use a public big data service for analysing the customer's own big data
Figure (optional)	
Pre-conditions (optional)	The big data services offered by the BDSP are registered in a public service catalogue.
Post-conditions (optional)	
Derived requirement	<ul style="list-style-type: none"> – Data delivery (refer to clause 8.3 requirement 1) – Customer support (refer to clause 8.4 requirements 1, 2, 3) – Personal information management (refer to clause 8.7 requirement 1)

II.2.2 Searching and utilizing a big data service

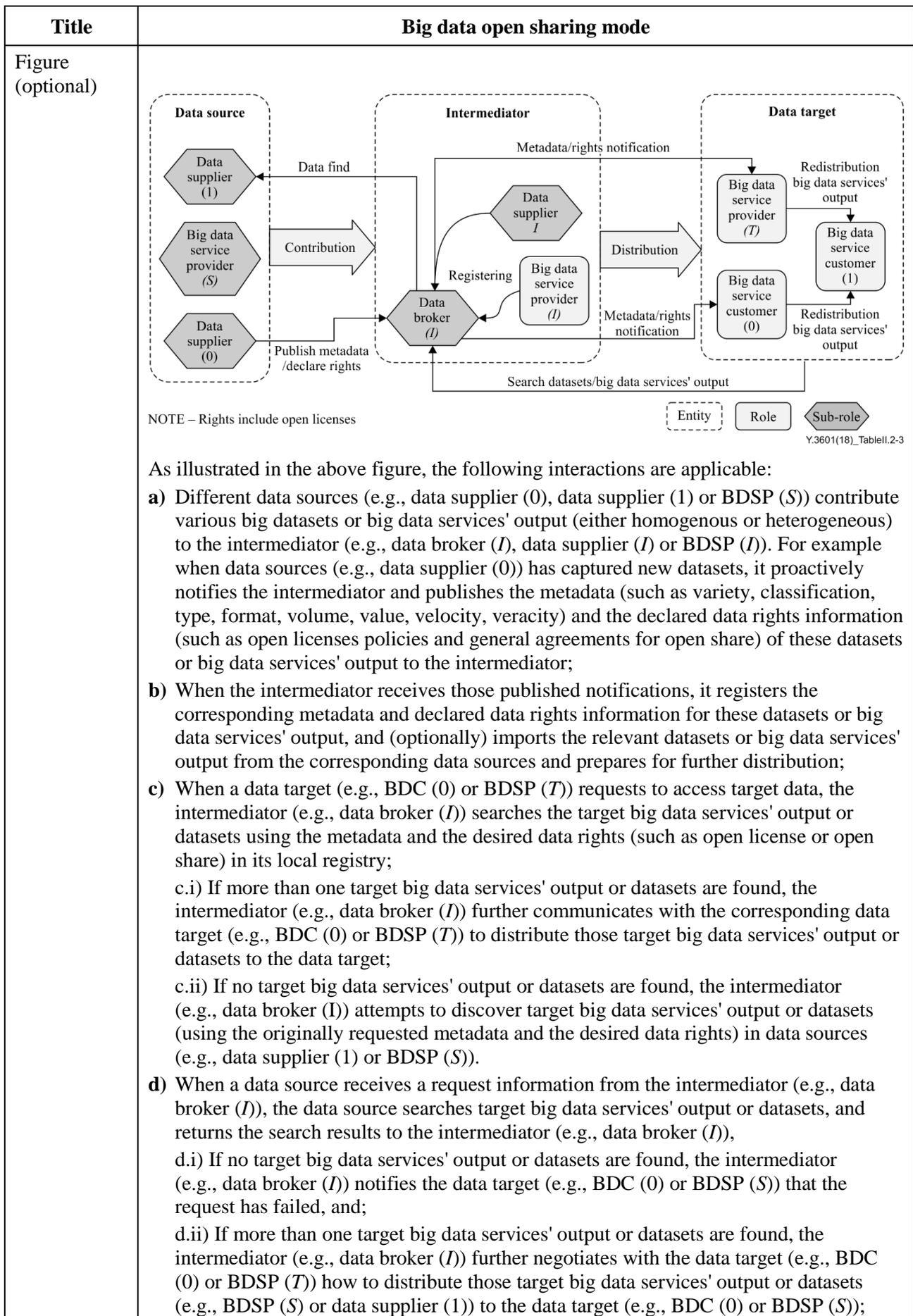
Title	Searching and utilizing a big data service
Description	A company (BDSP B) wants to use big data for their marketing purpose. An online data market (BDSP A + DP:data broker) is offering various data with different types and formats. A person in charge of marketing in the company finds data that suits its needs from the data market and downloads the corresponding data.
Roles/Sub roles	<ul style="list-style-type: none"> – DP:data supplier – DP:data broker – BDSP
Figure (optional)	

Title	Searching and utilizing a big data service
Pre-conditions (optional)	<p>Big data service provider A is storing various kinds of data. Big data service provider A imports data from data suppliers. Big data service provider A publishes data specifications in a data broker.</p> <p>NOTE 1 – A data specification is meta-information about the data which can be manipulated from static or streaming data in an on-demand manner (e.g., triggered when a user is requesting data).</p> <p>NOTE 2 – A data specification includes real-time data material with data access information, collection information, and information about a form in which data are stored, etc. In this case, BDSP connects DP:data supplier and requests & stores data for assembling data.</p>
Post-conditions (optional)	Big data service provider B provides an evaluation feedback about the data to data broker.
Derived requirement	<ul style="list-style-type: none"> – Data registration and cataloguing (refer to clause 8.1 requirements 7, 8) – Data retrieval (refer to clause 8.2 requirements 1, 3) – Data delivery (refer to clause 8.3 requirements 1, 3, 6) – Data quality management (refer to clause 8.5 requirements 1, 2) – Personal information management (refer to clause 8.7 requirement 2)

II.2.3 Big data sharing

II.2.3.1 Big data open sharing mode

Title	Big data open sharing mode
Description	<p>The various big datasets (either homogenous or heterogeneous) are contributed by different data sources (e.g., data providers, including data suppliers) to the intermediary (e.g., big data service provider and data broker), distributed by the intermediary to data targets (e.g., big data service customers) and accessed or redistributed by data targets without the protections (e.g., by copyright, patents, licenses or other restrictions of control mechanisms) and charge of payment. The big datasets are shared without the protections (e.g., by copyright, patents, licenses or other restrictions of control mechanisms) among the entities (e.g., data sources, intermediary and data targets), without charge of payment.</p> <p>NOTE – It might be protected by open license which is free of charge.</p>
Roles/Sub roles	<ul style="list-style-type: none"> – Data provider (DP) – DP:Data supplier – DP:Data broker – Big data service provider (BDSP) – Big data service customer (BDC)



Title	Big data open sharing mode
	<p>e) When the data target (e.g., BDC (0) or BDSP (T)) receives the negotiation information, it checks whether the target big data services' output or datasets should be accessed by itself or should be redistributed to its partners, and:</p> <p>e.i) If accessed by the data target (e.g., BDC (0) or BDSP (T)), then the data target (e.g., BDC (0) or BDSP (T)) requests the intermediary (e.g., data supplier (I) or BDSP (I)) to distribute the selected big data services' output or datasets to it for access (i.e., use);</p> <p>e.ii) if redistributed by the data target (e.g., BDC (0) or BDSP (T)), then the data target (e.g., BDC (0) or BDSP (T)) requests the intermediary (e.g., data supplier (I) or BDSP (I)) to receive the selected big data services' output in order to further redistribute the selected or produced big data services' output to its target partners (e.g., BDC (1)); or, the data target (e.g., BDC (0) or BDSP (T)) requests the intermediary (e.g., data supplier (I) or BDSP (I)) to directly distribute the selected or produced big data services' output to its partners (e.g., BDC(1)).</p>
Pre-conditions (optional)	<p>a) There is an agreement in place between the data source (e.g., data provider, including data supplier) and the intermediary (e.g., big data service provider and data broker);</p> <p>b) There is an agreement in place between the data target (e.g., big data service customer) and the intermediary (e.g., big data service provider and data broker);</p> <p>c) The data source (e.g., data provider, including data supplier) is responsible for the integrity, availability and authenticity of the datasets;</p> <p>d) The data source (e.g., data provider, including data supplier) published the metadata and declared data rights (i.e., open license policy, open share agreement);</p> <p>e) The data target (e.g., big data service customer) discovered the metadata and declared data rights (i.e., open license policy, open share agreement).</p>
Post-conditions (optional)	<p>a) The sharing mode related transactions between the involved entities (data sources, intermediators, data targets) and the resulting shared datasets or big data services' output should be publicly published online and archived for auditing offline;</p> <p>b) The sharing mode related transactions between the involved entities (data sources, intermediators, data targets) and the shared datasets or big data services' output were not repudiated.</p>
Derived requirement	<ul style="list-style-type: none"> – Data registration and cataloguing (refer to clause 8.1 requirements 2, 4) – Data retrieval (refer to clause 8.2 requirement 2) – Data delivery (refer to clause 8.3 requirement 7) – Data rights management (refer to clause 8.6 requirements 3, 4, 5)

II.2.3.2 Big data close sharing with/without charge modes

Title	Big data close sharing with/without charge modes
Description	<p>The various big datasets (either homogenous or heterogeneous) are contributed by different data sources (e.g., data provider, including data supplier) to the intermediary (e.g., big data service provider), distributed by the intermediary to data targets (e.g., big data service customers) and accessed or redistributed by data targets with the protection (e.g., by copyright, patents, licenses or other restrictions of control mechanisms) among the entities (e.g., data sources, intermediary and data targets), with or without charge of payment.</p> <p>NOTE – If the big datasets are shared with the protections of business trade agreement (e.g., contract), copyright, patents, licenses or other restrictions of control mechanisms, among the entities (e.g., data sources, intermediary and data targets), with charge of payment, it is exactly big data transaction use case.</p>

Title	Big data close sharing with/without charge modes
Roles/Sub roles	<ul style="list-style-type: none"> – Data provider (DP); – DP:data supplier; – DP:data broker; – Big data service provider (BDSP); – Big data service customer (BDC).
Figure (optional)	<p>NOTE – Rights include copyrights, patents and licenses policies.</p> <p style="text-align: right;">Y.3601(18)_TableII.2-4</p> <p>As illustrated in the above figure, the following interactions occur:</p> <ol style="list-style-type: none"> a) The different data sources (e.g., data supplier (0), data supplier (1) or BDSP (S)) contribute various big datasets or big data services' output (either homogenous or heterogeneous) to the intermediary (e.g., data broker (I), data supplier (I) or BDSP (I)). When a data source (e.g., data supplier (0)) has captured new datasets, it proactively notifies the intermediary and publishes metadata (such as variety, classification, type, format, volume, value, velocity, veracity) and the declared data rights information (such as copyrights, patents, licenses policies and price) of these datasets or big data services' output to the intermediary; b) When the intermediary receives notifications regarding published datasets, it registers the corresponding metadata and declared data rights information for these datasets or big data services' output, and (optionally) imports the relevant datasets or big data services' output from the corresponding data sources and prepares for further distribution; c) When a data target (e.g., BDC (0) or BDSP (T)) requests to access target data, the intermediary (e.g., data broker (I)) searches the target big data services' output or datasets using the metadata and the desired data rights in its local registry, <ol style="list-style-type: none"> c.i) If more than one target big data services' output or datasets are found, the intermediary (e.g., data broker (I)) further communicates with the corresponding data target (e.g., BDC (0) or BDSP(T)) to distribute those target big data services' output or datasets to the data target. c.ii) If none target big data services' output or datasets are found, the intermediary (e.g., data broker (I)) attempts to discover the target big data services' output or datasets (using the originally requested metadata and the desired data rights) in data sources (e.g., data supplier (1) or BDSP (S)); d) When a data sources receives a request information from the intermediary (e.g., data broker (I)), the data source searches target big data services' output or datasets, and returns the search results to the intermediary (e.g., data broker (I)), <ol style="list-style-type: none"> d.i) If no target big data services' output or datasets are found, the intermediary (e.g., data broker (I)) notifies the data target (e.g., BDC (0)) that the request has failed, and;

Title	Big data close sharing with/without charge modes
	<p>d.ii) If more than one target big data services' output or datasets are found, the intermediary (e.g., data broker (<i>I</i>)) further negotiates with the data target (e.g., BDC (0)) how to distribute those target big data services' output or datasets (e.g., BDSP (<i>S</i>) or data supplier (1)) to the data target (e.g., BDC (0) or BDSP (<i>T</i>));</p> <p>e) When the data target (e.g., BDC (0) or BDSP (<i>T</i>)) receives the negotiation information, it checks whether the target big data services' output or datasets should be accessed by itself or should be redistributed to its partners, and:</p> <p>e.i) If accessed by the data target (e.g., BDC (0) or BDSP (<i>T</i>)), then the data target (e.g., BDC (0) or BDSP (<i>T</i>)) requests the intermediary (e.g., data supplier (<i>I</i>) or BDSP (<i>I</i>)) to distribute the selected big data serves or datasets to it for access (i.e., use),</p> <p>e.ii) If redistributed by the data target (e.g., BDC(0) or BDSP (<i>T</i>)), then the data target (e.g., BDC (0) or BDSP (<i>T</i>)) requests the intermediary (e.g., data supplier (<i>I</i>) or BDSP (<i>I</i>)) to receive the selected big data services' output or datasets in order to further redistribute the selected or produced big data services' output to its target partners with protection of data rights; or, the data target (e.g., BDC (0) or BDSP (<i>T</i>)) requests the intermediary (e.g., Data supplier (<i>I</i>) or BDSP(<i>I</i>)) to directly distribute the selected or produced big data services' output to its partners (e.g., BDC(1)) for access (i.e., use) with protection of data rights.</p>
Pre-conditions (optional)	<p>a) There is an agreement in place between the data source (e.g., data provider, including data supplier) and the intermediary (e.g., big data service provider and data broker);</p> <p>b) There is an agreement in place between the data target (e.g., big data service customer) and the intermediary (e.g., big data service provider and data broker);</p> <p>c) The data source is responsible for the integrity, availability and authenticity of the datasets,</p> <p>d) The data source published the metadata and declared data rights (i.e., copyrights, patents, licenses policies and pricing),</p> <p>e) The data target (e.g., big data service customer) discovered the metadata and declared data rights (i.e., copyrights, patents, licenses policies and pricing).</p>
Post-conditions (optional)	<p>a) The sharing mode related transactions between the involved entities (data source, intermediary, data target) and the shared datasets or big data services' output should be publicly published online and archived for auditing offline,</p> <p>b) The sharing mode related transactions between the involved entities (data source, intermediary, data target) and the shared datasets or big data services' output were not repudiated.</p>
Derived requirement	– Data rights management (refer to clause 8.6 requirement 3).

II.2.4 Big data transaction

II.2.4.1 Big data transaction with ownership transfer mode

Title	Big data transaction with ownership transfer mode
Description	<p>Big datasets are exchanged between data sources (e.g., data providers, including data suppliers) and data targets (e.g., big data service customers) under the protection of business trade agreement, copyright, patents, licenses or other restrictions of control mechanisms, with charge of payment. The ownership of big datasets is transferred from data source to data target, according to the business trade agreement and licensing policy, the recipient of big datasets will have not only the right to access but also the right of disposal (such as redistribution) on the datasets.</p>

Title	Big data transaction with ownership transfer mode
Roles/Sub roles	<ul style="list-style-type: none"> – Data provider (DP) – DP: data supplier – DP: data broker – Big data service provider (BDSP) – Big data service customer (BDC)
Figure (optional)	<p>NOTE – Rights include copyrights, patents and licenses policies.</p> <p>Y.3601(18)_TableII.2-5</p> <p>As illustrated in the above figure, the following interactions are applicable:</p> <ol style="list-style-type: none"> a) A data source (e.g., data supplier (0), data supplier (1) or BDSP (S)) contributes new datasets or big data services' output and publishes metadata to the intermediary (e.g., data broker (I), data supplier (I) or BDSP (I)), with the declared data rights (i.e., copyrights, patents, licenses policies) and pricing policies; b) The intermediary receives datasets or big data services' output, registers metadata and re-publishes the metadata to a data target (e.g., BDC (0), BDC (1) or BDSP (T)); c) If a data target discovers the target datasets or big data services' output and accepts the declared data rights (i.e., copyrights, patents, licenses policies and pricing policies), the data target sends a purchase order to the intermediary; d) When the intermediary receives the purchase order from the data target, the intermediary distributes the required datasets to the data target, or produces the required big data services' output and provides to the data target; e) When the data target finishes the reception of datasets or big data services' output (e.g., by download or media delivery), the data target requests the data source and the intermediary (optionally) to delete the original datasets or the produced big data services' output, and the ownership of datasets will be transferred to the data target. NOTE – In this sense, the data target becomes a new data ownership that holds the data rights, and potentially, becomes another new data source.
Pre-conditions (optional)	<ol style="list-style-type: none"> a) There is a contract in place between the data source (e.g., data provider, including data supplier) and the intermediary (e.g., big data service provider and data broker); b) There is a contract in place between the data target (e.g., big data service customer) and the intermediary; c) The intermediary is responsible for matching the purchase and sale contract between the data source and the data target; d) The data source published the metadata and declared the copyrights, patents, licenses policies and pricing policies and guaranteed the integrity, availability and authenticity of the datasets; e) The data target discovered the metadata and accepted the declared copyrights, patents, licenses policies and pricing;

Title	Big data transaction with ownership transfer mode
	f) The data source, the intermediary and the data target complied with the business trade agreement, and the intermediary matched the purchase and sale contracts when appropriate between the data source and the data target.
Post-conditions (optional)	a) The dealing mode related transactions between the involved entities (data source, intermediary, data target) and the dealt datasets or big data services' output should be publicly published online and archived for monitoring online and auditing offline; b) The dealing mode related transactions between the involved entities (data source, intermediary, data target) and the dealt datasets or big data services' output were not repudiated; c) The ownership was transferred from the data source to the data target.
Derived requirement	– Data retrieval (refer to clause 8.2 requirement 4) – Data rights management (refer to clause 8.6 requirements 1, 2, 3)

II.2.4.2 Big data transaction without ownership transfer mode

Title	Big data transaction without ownership transfer mode
Description	Big datasets are exchanged between data sources (e.g., data providers, including data suppliers) and data targets (e.g., big data service customers) under the protections of business trade agreement, copyright, patents, licenses or other restrictions of control mechanisms, with charge of payment. The ownership of big datasets is unchanged according to the business trade agreement and licensing policy, the recipient of big datasets will have only the right to access but not the right of disposal (such as redistribution) on the datasets.
Roles/Sub roles	– Data provider (DP) – DP: data supplier – DP: data broker – Big data service provider (BDSP) – Big data service customer (BDC)
Figure (optional)	<p>NOTE – Rights include copyrights, patents and licenses policies.</p> <p>Y.3601(18)_TableII.2-6</p> <p>As illustrated in the above figure, the following interactions are applicable:</p> <ol style="list-style-type: none"> A data source (e.g., data supplier(0), data supplier(1) or BDSP(S)) contributes new datasets or big data services' output and publishes metadata to the intermediary (e.g., data broker (I), data supplier (I) or BDSP (I)), with the declared data rights (i.e., copyrights, patents, licenses policies) and pricing policies; The intermediary receives datasets or big data services' output, registers metadata and re-publishes the metadata to a data target (e.g., BDC (0), BDC (1) or BDSP (T));

Title	Big data transaction without ownership transfer mode
	<p>c) If a data target discovers the target datasets or big data services' output and accepts the declared data rights (i.e., copyrights, patents, licenses policies and pricing policies), the data target sends a purchase order to the intermediary;</p> <p>d) When the intermediary receives the purchase order from the data target, the intermediary distributes the required datasets to the data target, or produces the required big data services' output and provides to the data target;</p> <p>e) The data target receives datasets or big data services' output (e.g., by download or media delivery) and accesses the datasets or big data services' output. NOTE – In this case, data target will not hold the data rights, and consequently, is forbidden to redistribute the datasets to other data user (NOTE – None data target) or even data target.</p>
Pre-conditions (optional)	<p>a) There is a contract in place between the data source (e.g., data provider, including data supplier) and the intermediary (e.g., big data service provider and data broker);</p> <p>b) There is a contract in place between the data target(e.g., big data service customer) and the intermediary;</p> <p>c) The intermediary is responsible for matching the purchase and sale contract between the data source and the data target;</p> <p>d) The data source published the metadata and declared the copyrights, patents, licenses policies and pricing policies and guaranteed the integrity, availability and authenticity of the datasets;</p> <p>e) The data target discovered the metadata and accepted the declared copyrights, patents, licenses policies and pricing policies;</p> <p>f) The data source, the intermediary and the data target complied with the business trade agreement, and the intermediary matches the purchase and sale contracts when appropriate between the data source and the data target.</p>
Post-conditions (optional)	<p>a) The dealing mode related transactions between the involved entities (data source, intermediary, data target) and the dealt datasets or big data services' output should be publicly published online and archived for monitoring online and auditing offline;</p> <p>b) The dealing mode related transactions between the involved entities (data source, intermediary, data target) and the dealt datasets or big data services' output were not repudiated;</p> <p>c) The ownership was not transferred from the data source to the data target and it was still kept by the data source.</p>
Derived requirement	– Data rights management (refer to clause 8.6 requirement 3)

II.2.5 Big data interconnection

II.2.5.1 Big data interconnection peering mode

Title	Big data interconnection peering mode
Description	Big data interconnection is a model of big data exchange for enabling the interworking between two or more data sources (e.g., data providers) for mutual exchange of big datasets, by which the source big datasets are interchanged among data sources (e.g., data providers), contributed to the intermediary (e.g., big data service provider) and distributed to data targets (e.g., big data service customers).
Roles/Sub roles	<ul style="list-style-type: none"> – Data provider (DP) – DP:data supplier – DP:data broker – Big data service provider (BDSP) – Big data service customer (BDC)

Title	Big data interconnection peering mode
Figure (optional)	<p>The diagram illustrates the 'Big data interconnection peering mode'. It features three 'Data source' entities, each containing a 'Data supplier' (represented by a hexagon) and a 'Big data service provider' (represented by a rounded rectangle). Source 1 (top left) has a 'Data supplier (1)' and 'Big data service provider (1)'. Source 2 (bottom left) has a 'Data supplier (2)' and 'Big data service provider (2)'. Source 0 (center right) has a 'Data supplier (0)' and 'Big data service provider (0)'. Bidirectional arrows labeled 'datasets or big data services' output' connect sources 1 and 2 to source 0. A large arrow labeled 'Contribution' points from source 0 to an 'Intermediator' entity. From the 'Intermediator', an arrow labeled 'Distribution' points to 'Data targets'. A legend at the bottom right identifies 'Entity' (dashed box), 'Role' (rounded rectangle), and 'Sub-role' (hexagon). Reference: Y.3601(18)_TableII.2-7</p> <p>As illustrated in the above figure, the following interactions are applicable:</p> <ol style="list-style-type: none"> Three data sources interchange directly with each other, and one of them is selected as primary data source (e.g., data supplier (0) BDSP (0)) as serving node to connect with the intermediary, and the selected primary data source contributes all datasets or big data services' output from data sources (e.g., both data supplier (0) BDSP (0), data supplier (1) BDSP (1) and data supplier (2) BDSP (2)) to the intermediary; One data source is selected as primary data source, the other data sources are selected as secondary data sources (e.g., data supplier (1) BDSP (1) and data supplier (2) BDSP (2)) by default. The secondary data sources interchange datasets or big data services' output with the primary data source on request in order to provide their datasets or big data services' output via the primary data source to the intermediary; The intermediary distributes the datasets or the big data services' output to the data targets on request.
Pre-conditions (optional)	<ol style="list-style-type: none"> There is a contract in place between the data sources (e.g., DP:data suppliers, DP:data brokers or BDSPs) and the intermediary; There is a contract in place between the data targets (e.g., BDCs) and the intermediary; The intermediary is responsible for matching the purchase and sale contract between the data sources and the data targets; The data sources published the metadata and declared the copyrights, patents, licenses policies and pricing policies and guaranteed the integrity, availability and authenticity of the datasets; The data targets discovered the metadata and accepted the declared copyrights, patents, licenses policies and pricing policies; The data sources, the intermediary and the data targets complied with the business trade agreement, and the intermediary matches the purchase and sale contracts when appropriate between the data sources and the data targets.
Post-conditions (optional)	<ol style="list-style-type: none"> The interconnecting mode related transactions between the involved entities (data sources, intermediary, data targets) and the interconnected datasets or big data services' output should be publicly published online and archived for monitoring online and auditing offline; The interconnecting mode related transactions between the involved entities (data sources, intermediary, data targets) and the interconnected datasets or big data services' output were not repudiated.

Title	Big data interconnection peering mode
Derived requirement	– Data delivery (refer to clause 8.3 requirement 5)

II.2.5.2 Big data interconnection federation mode

Title	Big data interconnection federation mode
Description	Big data interconnection is a model of big data exchange for enabling the interworking between two or more data sources (e.g., data providers) for mutual exchange of big datasets, by which the source big datasets are interchanged among data sources (e.g., data providers), contributed to the intermediary (e.g., big data service provider) and distributed to data targets (e.g., big data service customers).
Roles/Sub roles	<ul style="list-style-type: none"> – Data provider (DP) – DP:data supplier – DP:data broker – Big data service provider (BDSP) – Big data service customer (BDC)

Figure (optional)	<p>The diagram illustrates the Big data interconnection federation mode. It shows three data sources on the left, each containing a Data supplier (1, 2, or 0) and a Big data service provider (1, 2, or 0). Arrows labeled 'datasets or big data services' output' point from each data source to a central Data source (0) which also contains a Data supplier (0) and a Big data service provider (0). An arrow labeled 'Contribution' points from this central data source to an 'Intermediator'. An arrow labeled 'Distribution' points from the 'Intermediator' to 'Data targets'. A legend at the bottom right identifies Entity (dashed box), Role (rounded rectangle), and Sub-role (hexagon). Reference: Y.3601(18)_TableII.2-8</p> <p>As illustrated in the above figure, the following interactions are applicable:</p> <ol style="list-style-type: none"> a) A group of peering data sources (e.g., data supplier (0) BDSP (0), data supplier (1) BDSP (1) and data supplier (2) BDSP (2)) interchange directly with each other (e.g., data supplier (0) BDSP (0) with data supplier (1) BDSP (1) and data supplier (2) BDSP (2), data supplier (1) BDSP (1) with data supplier (0) BDSP (0) and data supplier (2) BDSP (2), and data supplier (2) BDSP (2) with data supplier (0) BDSP (0) and data supplier (1) BDSP (1)), and who mutually combine or integrate their datasets or big data services' output and provide to the intermediary; b) One data source (e.g., data supplier (0) BDSP (0)) is selected as serving node to connect with the intermediary and contributes all datasets or big data services' output from the data sources group (e.g., data supplier (0) BDSP (0), data supplier (1) BDSP (1) and data supplier (2) BDSP (2)) to the intermediary; c) The intermediary distributes the datasets or the big data services' output to the data targets (e.g., BDCs) on request.
Pre-conditions (optional)	<ol style="list-style-type: none"> a) There is a contract in place between the data sources (e.g., DP:data suppliers, DP:data brokers or BDSPs) and the intermediary; b) There is a contract in place between the data targets (e.g., BDCs) and the intermediary;

Title	Big data interconnection federation mode
	<ul style="list-style-type: none"> c) The intermediary is responsible for matching the purchase and sale contract between the data sources and the data targets; d) The data sources published the metadata and declared the copyrights, patents, licenses policies and pricing policies and guaranteed the integrity, availability and authenticity of the datasets; e) The data targets discovered the metadata and accepted the declared copyrights, patents, licenses policies and pricing policies; f) The data sources, the intermediary and the data targets complied with the business trade agreement, and the intermediary matches the purchase and sale contracts when appropriate between the data sources and the data targets.
Post-conditions (optional)	<ul style="list-style-type: none"> a) The interconnecting mode related transactions between the involved entities (data sources, intermediary, data targets) and the interconnected datasets or big data services' output should be publicly published online and archived for monitoring online and auditing offline; b) The interconnecting mode related transactions between the involved entities (data sources, intermediary, data targets) and the interconnected datasets or big data services' output were not repudiated.
Derived requirement	– Data delivery (refer to clause 8.3 requirement 5)

II.2.5.3 Big data interconnection intermediary mode

Title	Big data interconnection intermediary mode
Description	<p>Big data interconnection is a model of big data exchange for enabling the interworking between two or more data sources (e.g., data providers) for mutual exchange of big datasets, by which the source big datasets are interchanged among data sources (e.g., data providers), contributed to the intermediary (e.g., big data service provider) and distributed to data targets (e.g., big data service customers).</p> <p>In the data interconnection intermediary mode, a data source interworks with one or more peer data sources and provides intermediation, aggregation and arbitrage of datasets provided by these data sources.</p>
Roles/Sub roles	<ul style="list-style-type: none"> – Data provider (DP) – DP:data supplier – DP:data broker – Big data service provider (BDSP) – Big data service customer (BDC)

Title	Big data interconnection intermediary mode
Figure (optional)	<p>As illustrated in above figure, the following interactions are applicable:</p> <ol style="list-style-type: none"> One data source (e.g., data supplier (0) BDSP (0)) is selected as serving node to connect with the intermediary, and the selected data source contributes all datasets or big data services' output on behalf of the other peering data sources (e.g., data supplier (1) BDSP (1) and data supplier (2) BDSP (2)) to the intermediary; The selected data source (e.g., data supplier (0) BDSP (0)) interchanges directly with the other peering data sources (e.g., data supplier (0) BDSP (0) with data supplier (1) BDSP (1) and data supplier (2) BDSP (2)) and provides aggregation, integration and intermediation of their datasets or big data services' output to the intermediary; The intermediary distributes the datasets or the big data services' output to the data targets (e.g., BDCs) on request.
Pre-conditions (optional)	<ol style="list-style-type: none"> There is a contract in place between the data sources (e.g., DP:data suppliers, DP:data brokers or BDSPs) and the intermediary; There is a contract in place between the data targets (e.g., BDCs) and the intermediary; The intermediary is responsible for matching the purchase and sale contract between the data sources and the data targets; The data sources published the metadata and declared the copyrights, patents, licenses policies and pricing policies and guaranteed the integrity, availability and authenticity of the datasets; The data targets discovered the metadata and accepted the declared copyrights, patents, licenses policies and pricing policies; The data sources, the intermediary and the data targets complied with the business trade agreement, and the intermediary matches the purchase and sale contracts when appropriate between the data sources and the data targets.
Post-conditions (optional)	<ol style="list-style-type: none"> The interconnecting mode related transactions between the involved entities (data sources, intermediary, data targets) and the interconnected datasets or big data services' output should be publicly published online and archived for monitoring online and auditing offline; The interconnecting mode related transactions between the involved entities (data sources, intermediary, data targets) and the interconnected datasets or big data services' output were not repudiated.
Derived requirement	<ul style="list-style-type: none"> Data delivery (refer to clause 8.3 requirement 5)

Bibliography

- [b-ITU-T Y.2201] Recommendation ITU-T Y.2201 (2011), *Requirements and capabilities for ITU-T NGN*.
- [b-ITU-T Y.2701] Recommendation ITU-T Y.2701 (2007), *Security requirements for NGN release 1*.
- [b-ITU-T Y.3502] Recommendation ITU-T Y.3502 (2014), *Information technology – Cloud computing – Reference architecture*.
- [b-ISO/IEC 2382] ISO/IEC 2382 (2015), *Information technology – Vocabulary*.

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	Tariff and accounting principles and international telecommunication/ICT economic and policy issues
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Cable networks and transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Environment and ICTs, climate change, e-waste, energy efficiency; construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality, telephone installations, local line networks
Series Q	Switching and signalling, and associated measurements and tests
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	Global information infrastructure, Internet protocol aspects, next-generation networks, Internet of Things and smart cities
Series Z	Languages and general software aspects for telecommunication systems