

That's my DATA!

Keep control of your data!

BOMOD: A model to maintain and manage your data

BigTU

Silja Eckartz

Jack Verhoosel

2 December 2015

Background: Data in BigTU

- › Market information to better match supply and demand
- › Growers do not have information on where their products will be consumed and are unable to respond to specific consumer demands
- › Data important for horticulture (flowers/ plants and fruit/ vegetables)
 - › Production data
 - › Transaction data
 - › Export data
 - › Consumer data (demand)

Issues experienced in the horticulture sector related to data sharing

- › Concerns of entrepreneurs and information service providers which result in limited data sharing:
 - › The data is sensitive with respect to competitive relationships:
 - › Loss of advantage if market opportunities are rapidly revealed to competitors
 - › Loss of investments in private data collection systems
 - › Opening detailed information on the Dutch sector to international competitors
 - › Risk of negative publicity due to wrong interpretation of open data before the sector can properly respond
 - › Missing incentives and limited Reciprocity

Issues experienced in the horticulture sector related to data sharing

- › Government does not share the data that would be needed by participants to develop applications that are valuable for the sector
 - › Timeliness of data access → when is data shared
 - › Questions on ownership
 - › Sharing of competitive relevant information (Economic sensitivity)
 - › Quality of data → liability

Some first initiatives

- › Analysis of Twitter data (open data)
- › Use Cases in BigTU: WAPA
- › FloriData: Sharing of data related to payment (payment behaviour of customers)
- ›

(Open) Data – An Introduction

- › Data = new oil or new gold for innovation and economic growth
- › Open Data is data that is provided for unrestricted reuse. This is usually achieved through making it machine-readable, publishing it online, and providing a clear statement that gives anyone permission to reuse it. Anyone can use open data for anything, subject to other relevant laws and regulations relating to responsible data use.
- › Open data aims for organizations to become more transparent and thereby accountable to citizens
- › **Different degrees of “open” when sharing data**
 - › Closed data: data that should not be shared
 - › Shared data: data available for specific users or particular kinds of reuse
 - › Open data: data provided for unrestricted reuse

Data Governance – An Introduction

- › Data governance: covers aspects such as data quality, data management, data ownership, metadata management, access rights, decision rights, accountability, and data policies.

- › (Open) data decision tree
 - › Which data should be open/ closed, for whom?
 - › Which barriers can you expect?
 - › How to overcome these

Deciding which data to open

- Open Data Decision Tree: Method to decide which data sets are suitable to be published as “open” data

Step 1: Goals of sharing (open) data

Step 2: Incentives for sharing data

Step 3: Constraints to open data

Step 4: Process to open/ share data

Step 1: Goals of sharing (open) data

New insights

Improved
planning

Improved
decision making

Increase
transparency

Knowledge
sharing

Better Supply &
Demand
matching

Improved
services

Address societal
challenges

New Data
Applications

Empower sector

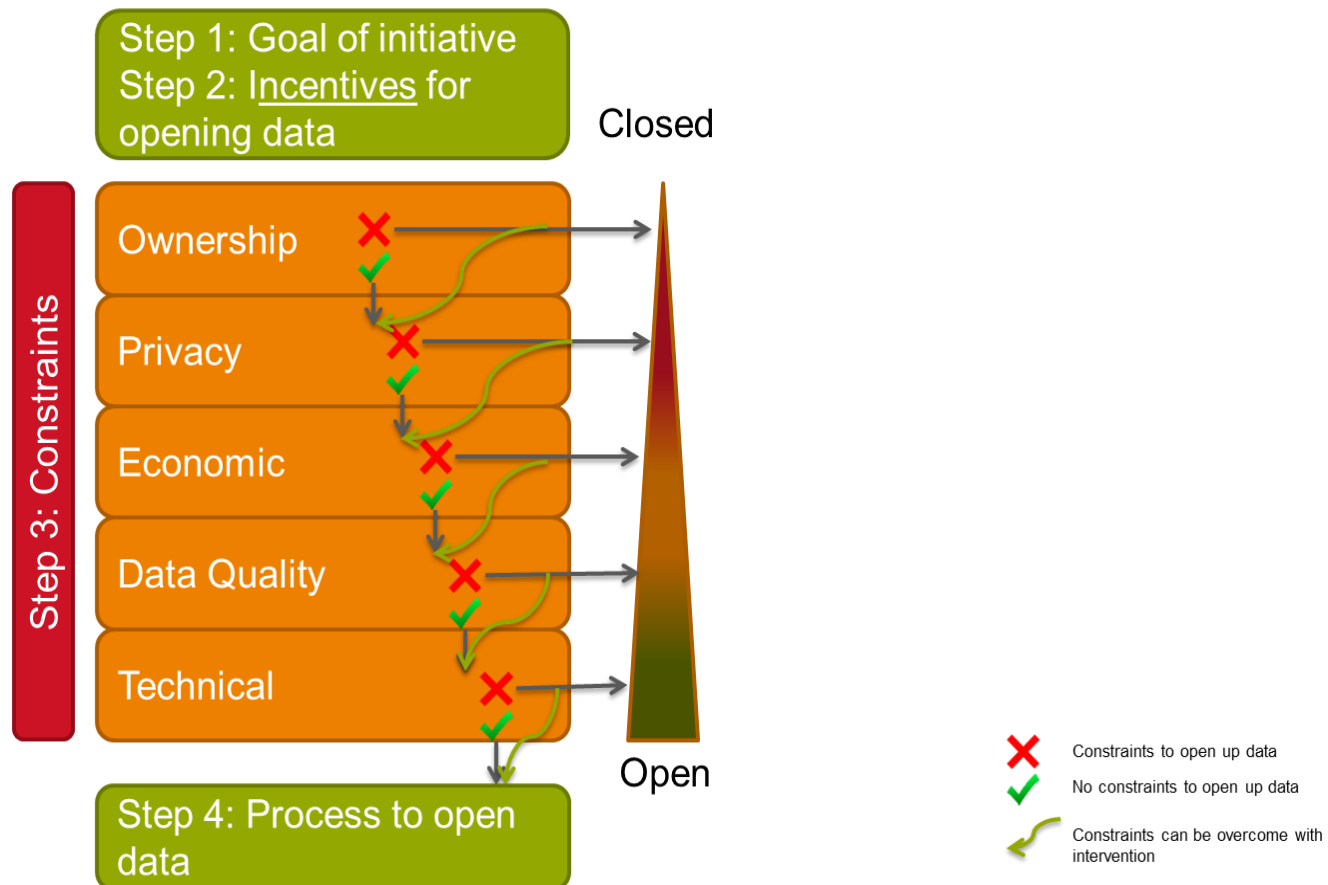
Data driven
Innovation

Improved
policies

Step 2: Incentives for sharing/ opening data

Economic Incentives	Legal and regulatory Incentives
<ul style="list-style-type: none">• Offer new services resulting in more sales• Efficiency gains; cost reductions• Creating a competitive advantage; influence market, or be perceived as a leader• Improved image resulting in more sales• Enlarge customer base• Collaboration to be able to make the market as a whole grow• Fear of losing a reputation or losing out on potential gains• Societal goals leading to economic advantages• Reciprocity: Open data in exchange for data from another party	<ul style="list-style-type: none">• Compliance to regulations, for example environmental, data protection, or safety regulations• Transparency, ensuring trust and accountability• Compliance to directives regarding open data

Step 3: Decision model to publish/ share individual datasets



Zoom-in on Constraints 1/3

- › **Ownership:** Only the owner of the data can make a decision about whether to publish data or not. If more than one organization owns data, all involved organizations need to agree on opening up the data. The owner is responsible for the quality and maintenance of the data.
 - › Interventions: Trust building, licensing, rights policy

- › **Privacy:** If a data set contains classified or privacy sensitive information that can be traced back to individual persons or companies this will constrain the data owner to publish the data. Legal constraints related to the privacy of data can also present a constraint to publishing data.
 - › Interventions: Anonymization, aggregation, access control

Zoom-in on Constraints 2/3

- › **Economic:** In the case that a data owner currently earns money by providing his data, this will constrain the publication of the data. Economic sensitivity may constrain the data to be opened up. High costs of opening data are a constraint.
 - › Interventions: Share costs, define pricing structure

- › **Data quality:** Metadata is needed describing the quality of the data, data with too little quality should rather not be published. Liability issues prohibit open data. Data gathered in one context might not be usable in another context.
 - › Interventions: Metadata, context information, data cleansing

Zoom-in on Constraints 3/3

- › **Technical:** If the data is unstructured it may be difficult to convert it into a machine-readable format relevant to a data user. The size of the data set, the existence of a semantic model, and identifiers are other technical issues that need to be considered.
- › **Interventions:** Offer data in a structured format, reuse existing vocabularies, publish data according to existing data standards

Result data decision tree: Who is allowed to see what part of the data

Who	What level
Public	Closed
Community	Semi-open (abstract level data)
Use case partners	Open data
...	...

Data Governance Dashboard

▼ Data Governance Dashboard - Welcome

Dataset selectie

User Group definitie

▼ Vragen over het gehele dataset

Economische sensitiviteit

Data Quality

▼ Vragen over specifieke datavelden

Privacy

Data Quality

Interventies

User Group X	Aggregated
No one	All data
All user groups	Anonymized
Everyone except user group X	Filtered

Hoe wilt u gebruik maken van onze data governance dashboard?

zonder registratie ▼

Als u al lid bent van de data platform kunt u hier inloggen:

Name:

Password:

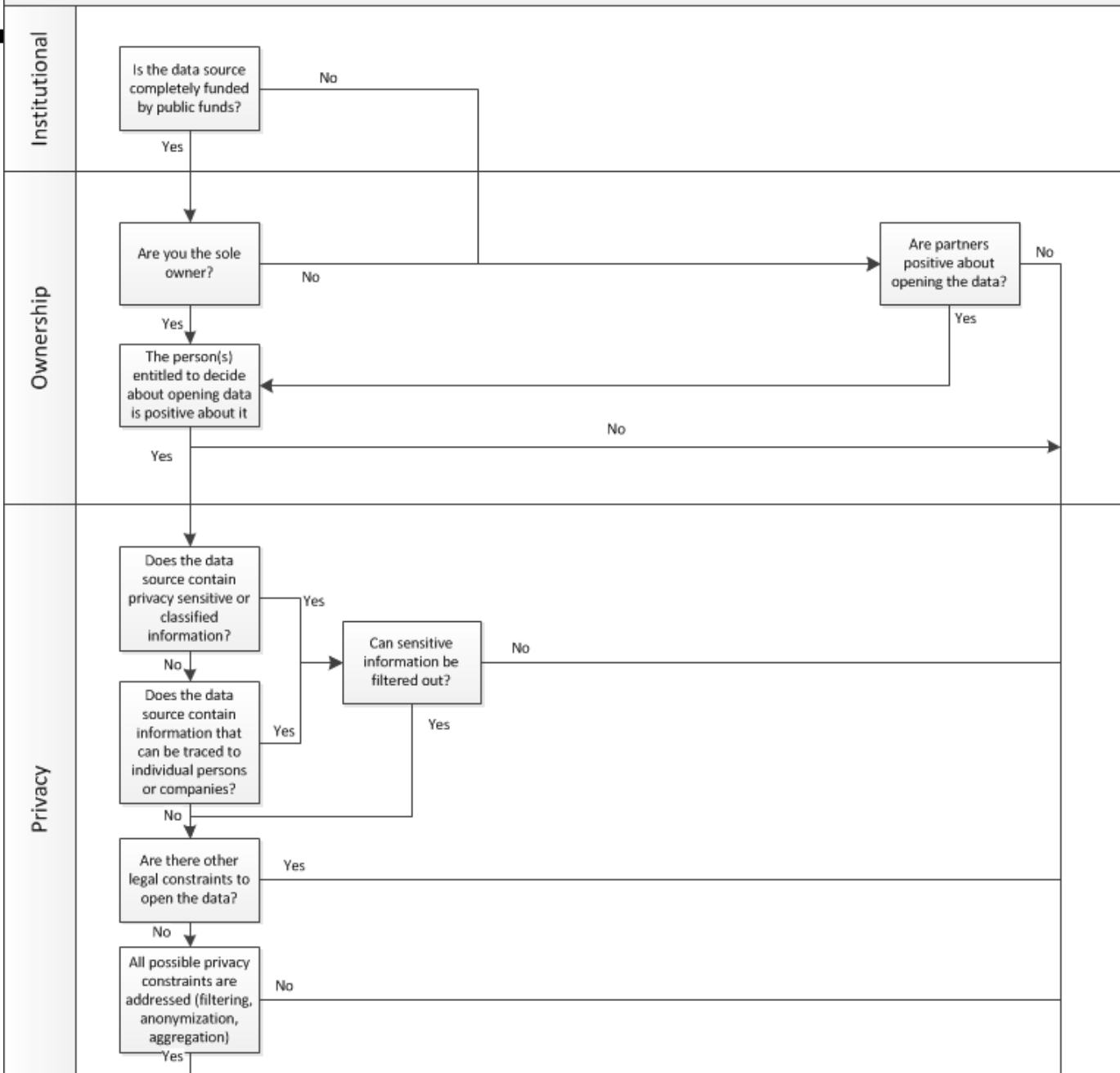
Als u zonder registratie een dataset wilt analyseren kunt uw hier clicken: [Dataset selectie](#)

→

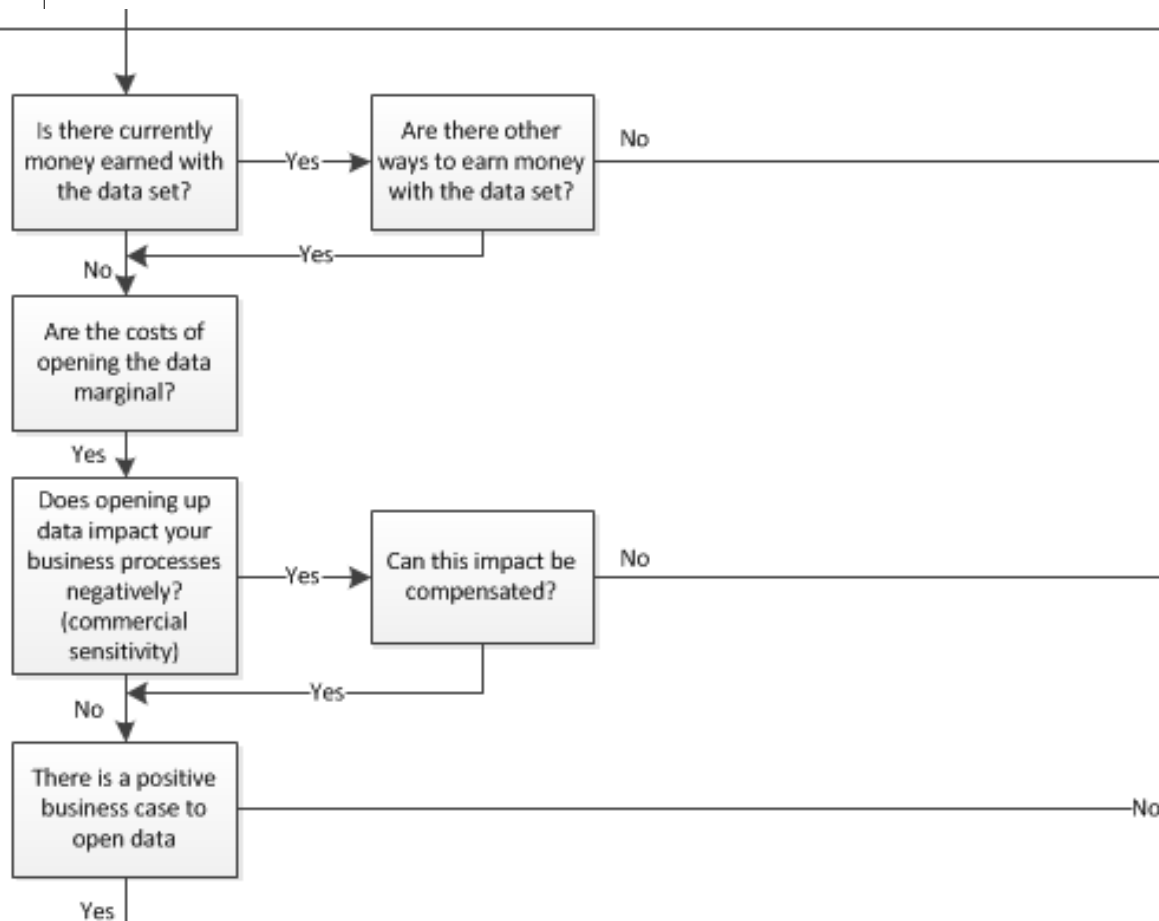
Step 3: Interventions

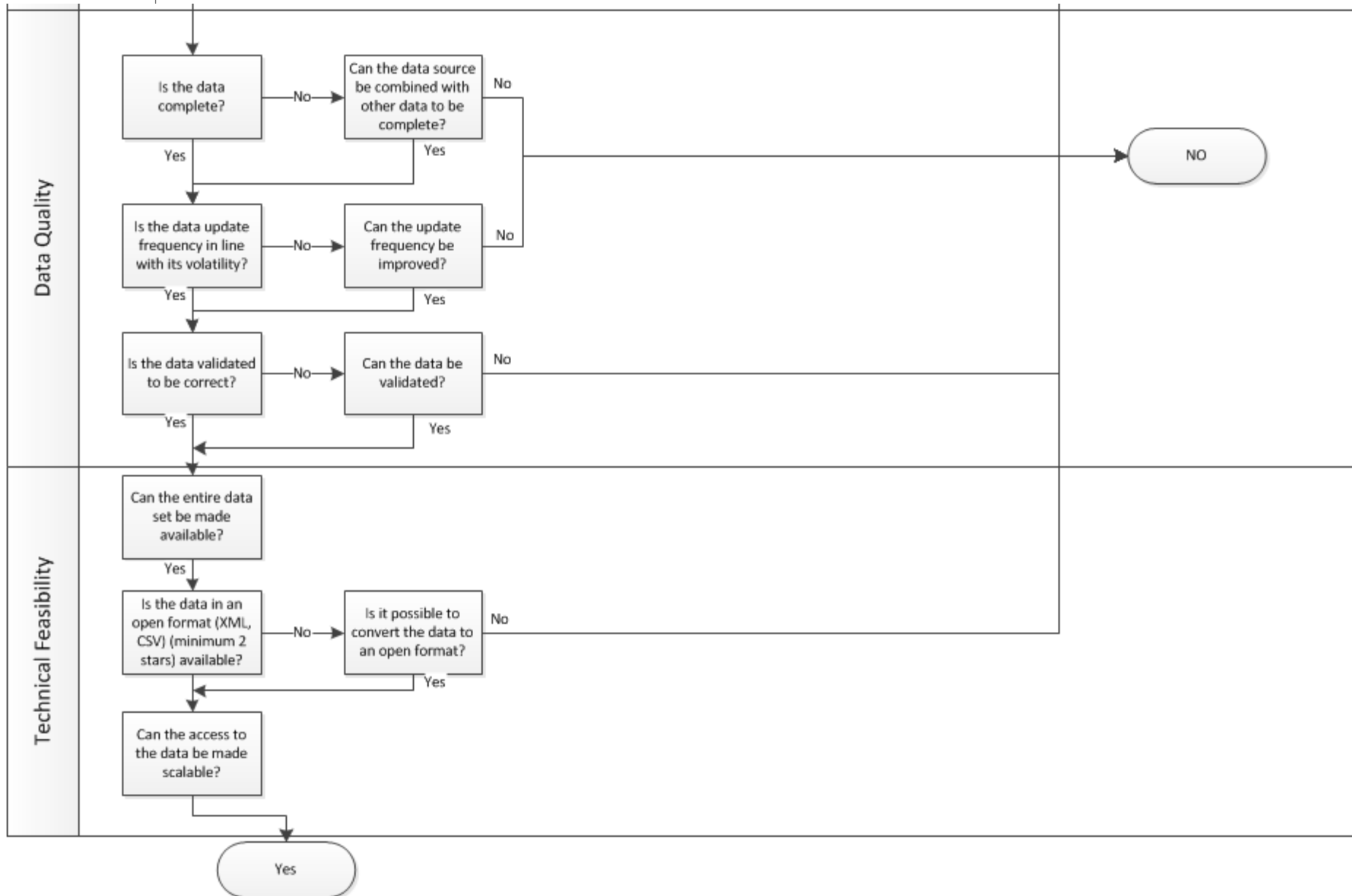
Technical intervention	Organizational intervention
<ul style="list-style-type: none">• Enabling aggregation or anonymization of data• Technical access control• Introduce yield management• Improve data quality of data set• Publish meta-data about the data set• Make data in more structured format available (5 star model)• Support existing data standards	<ul style="list-style-type: none">• Obtaining permission by data subjects to publish data• Collaboration mechanisms• Trust mechanisms• Successful pilot projects• Sharing of best practices• Cultural change program

Decision Model for Open Data



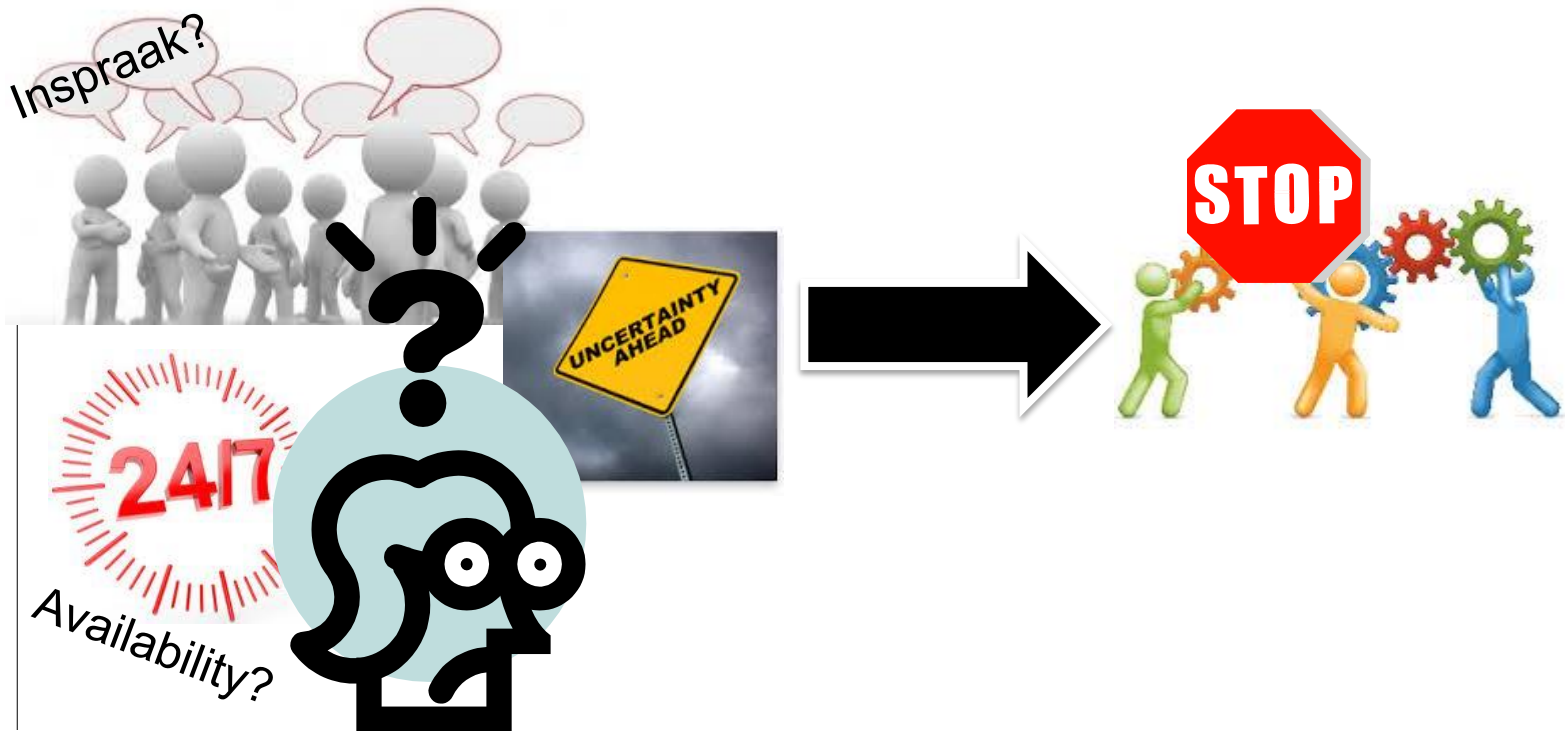
Economic





Data owners, do not forget! Maintenance of data is important!

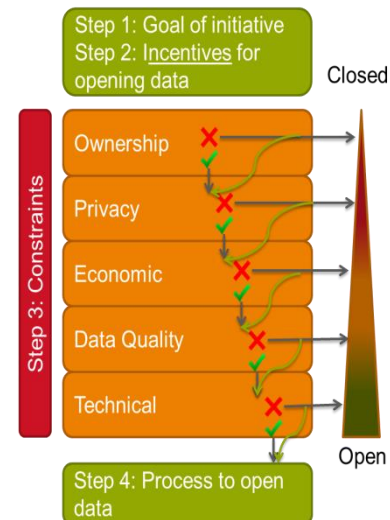
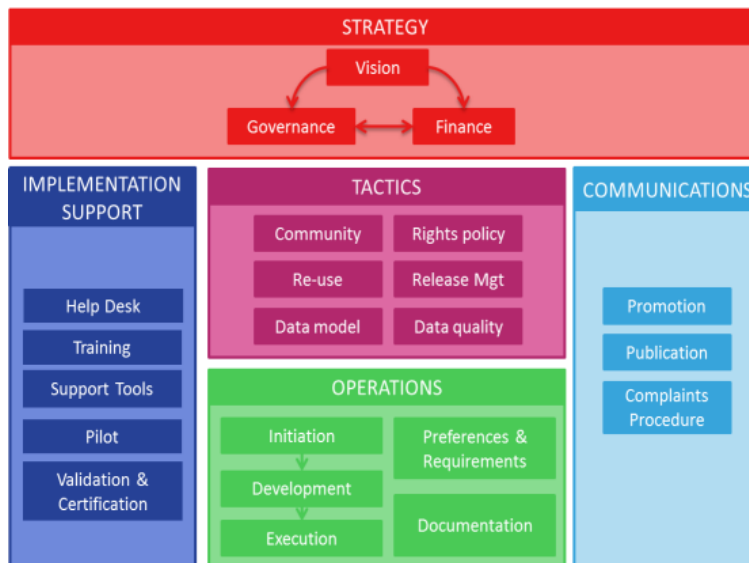
- › In the Netherlands open data is often not well maintained



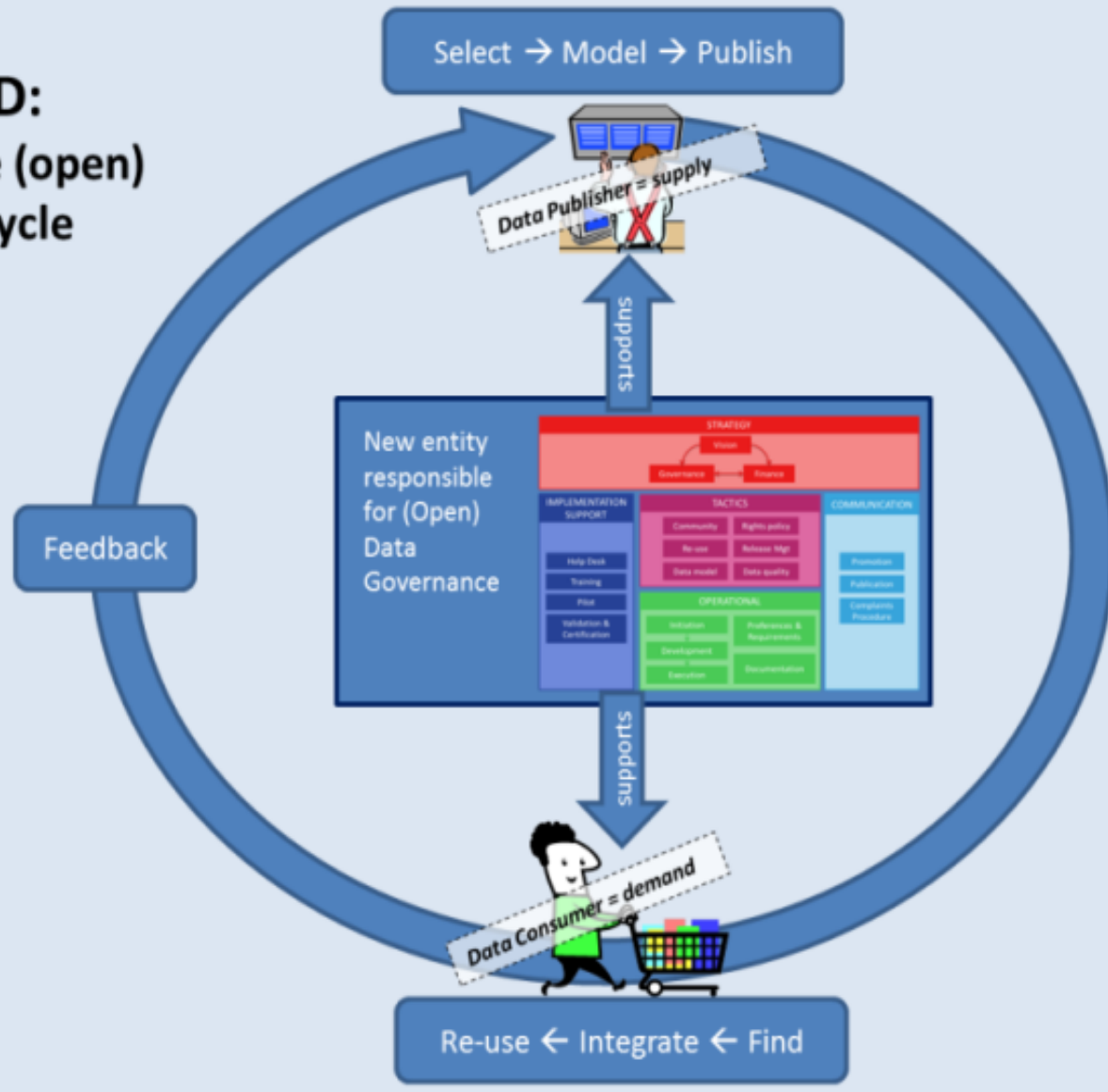
BOM(O)D

Method to manage and develop (open) data

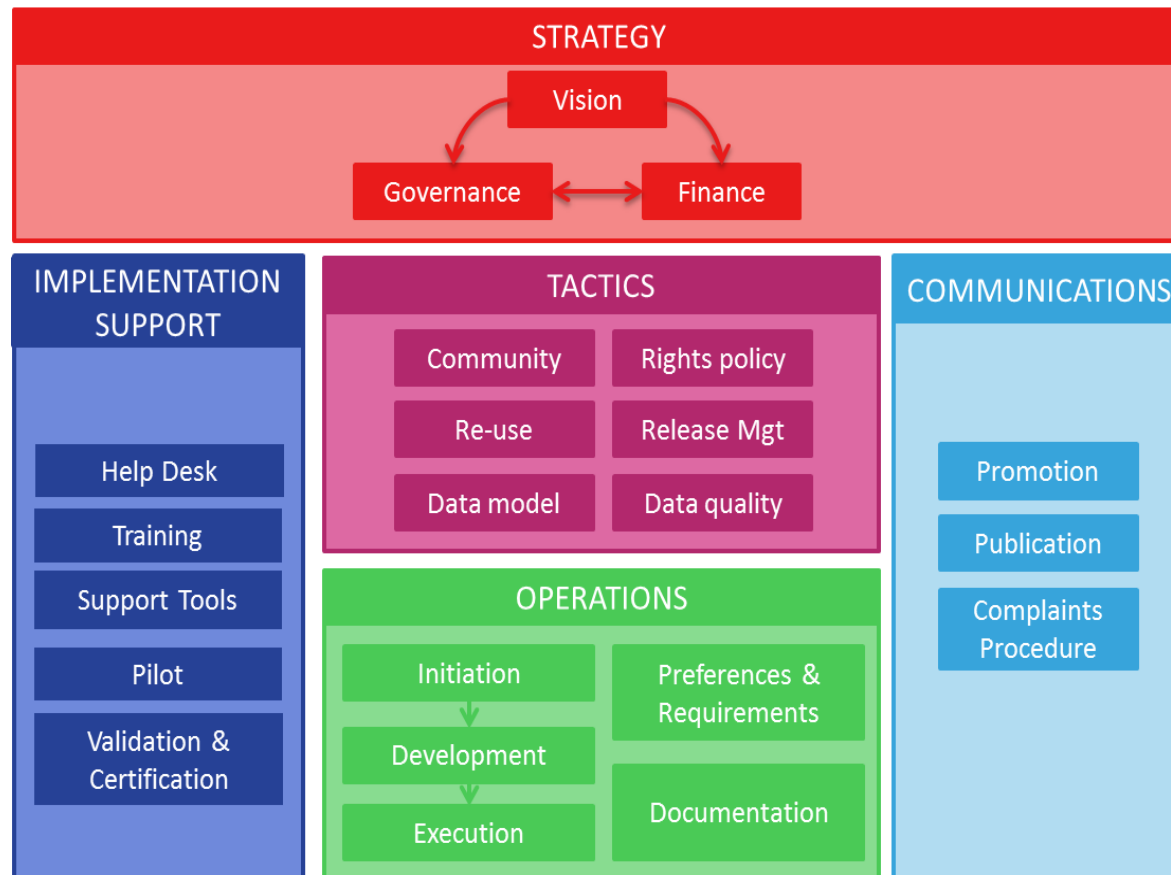
- Definition of a set of activities (incl. practical guidelines and tools) to manage and develop (open) data
 - Amongst others: Open Data Decision Tree: Method to decide which data sets are suitable to be published as “open” data
- Maturity model to i) decide with which activities to start and ii) measure the level of maturity of organizations managing open data



BOMOD: Managing the (open) Data Lifecycle



BOMOD – Activity Diagram



› **Strategy:**

- describes directing activities related to the strategic (long) term focus of the data governance entity including vision, governance structure and financial structure for sharing (open) data
- performed by policy makers/ higher management/ decision makers

› **Tactics:**

- describes six activities that steer the process of publishing/ sharing data on a tactical level
- performed in collaboration between policy makers and executing departments

› **Operations:**

- describes executive activities that lead to the actual publication/ sharing of new datasets
- performed by executing departments, such as IT departments

› **Implementation Support:**

- describes activities focusing on supporting the publication of data sets
- performed by the IT department and/ or help desk

› **Communication:**

- describes supporting activities focusing on creating support for publishing open data
- done in collaboration with the marketing department

Next Steps for BigTU

- › Develop overview of how data is flowing through the sector
- › Apply data governance (e.g. BOMOD) to ensure that concerns of entrepreneurs and ISPs are considered carefully
 - › Identify incentives for organizations to share data
 - › Convince sector that data can be shared safely via Horticulture

Questions & Discussion