A FAKE STORY IN A TRUSTWORTHY GUIDE TO THE FAIR PRINCIPLES FOR RESEARCH DATA.

A FAIRy tale

Published by: FAIR project

Text written by:

Karsten Kryger Hansen ORCID 0000-0002-2407-8764

Mareike Buss ORCID 0000-0002-1459-1345

Lea Sztuk Haahr ORCID 0000-0003-2211-6138

Editorial Design and Illustrations by: Paulina Halina Sieminska

Proofreading : Lotte Stehouwer Øgaard

FAIR project is realised within cooperation of following Institutions: DeiC / Deff / DTU / CBS / AAU / KU / Rigsarkivet / Det Kgl Bibliotek

Error

Page not found!

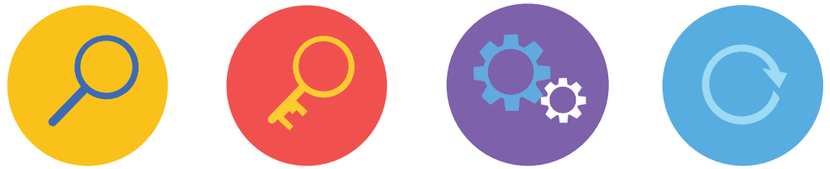
Thanks for input and support to everyone involved in the project.
Thanks to DM Forum and its sponsors for making this possible.

Table of contents:

Introduction

Once upon a time in the beautiful kingdom of Datamania lived a prince named Prince Fairhair. Though he was gentle as few, and good looking too, his father would not let him choose the love of his life on his own. No, he was destined to marry a woman from the neighbouring kingdom. He did not even know her name, only that she was referred to as My Fair Lady. Before the father of My Fair Lady could accept the marriage, he had a quest for Prince Fairhair. Only by fulfilling the quest, would he be able to marry the princess. His quest was to find out how to turn water into gold. A quest that would require gathering loads of data chests and look for clues that could lead to the recipe.

Luckily, Prince Fairhair was not alone in his quest. One of the castle wings housed a number of wizards who could help him decrypt and investigate the data chests. However, it was impossible for the data wizards to go and hunt for data themselves. Thus to assist them, a huge number of elves were trained to look for data chests. The elves had read books, journals, comics and even poetry to know where to look. The quest was about to begin, and the elves went hunting for data chests all over the kingdom of Datamania and in empires far far away.



The FAIR principles were first published in 2016. They contain guidelines for good data management practice that aim at making data FAIR: Findable, Accessible, Interoperable and Reusable. Each letter refers to a list of principles with a total of 15 principles altogether.

Although they originate from the life sciences, the principles can be – and have been - applied within other research disciplines, including the humanities and the social sciences. Since their publication, the European Union as well as individual funders and universities have declared their support for and approval of the FAIR principles. This spans from the creation of data management tools and infrastructures to defining policies for data handling. Some implementations stick closely to the original definitions, while others are inspired by the spirit of the FAIR principles.

A fundamental prerequisite for understanding FAIR is to know that both humans and machines are intended as digesters of data. This will lead to an ecosystem that is fast to respond to change and automatically adapts to new findings or changes. That is the reason for focusing on standards for the data, identification mechanisms, availability of data etc. Secondly, the FAIR principles apply to both data and its metadata – i.e. records about data sets. That is why the term "(meta)data" is stated in the principles. Thirdly, the principles are not only about open data. You can work in a *FAIR manner* with data that is not intended for public availability.

The FAIR principles do not represent a quality standard that you can use to evaluate tools, data, policies etc. This would soon make the principles out-of-date and inapplicable across research disciplines. The implementation of FAIR can be a gradual and systematic adaptation of new work routines or a huge leap where you replace one type of infrastructure with another. The implementation of the principles should be adapted to each research area, meaning that each community will make the principles work in their respective contexts.

Findable #1:

(Meta)data are assigned globally unique and persistent identifiers

The elves returned one by one to the castle, and some of them were really frustrated. They had been following paths to data chests that had been meticulously described, but somehow the data chests had been removed, just leaving holes in the ground. Fimble was one of these elves, who came back quite puzzled about some strange codes he had found. He could not decipher them and therefore did not know where to go.

"*Look*" said Fimble to the data wizard, "*I have this strange code 10.1234/abbb, and I don't know what it means?*"

"*Oh, these are very useful indeed*" said the data wizard. *"We can look up the codes in these huge books. Now let me see. 10 is the great country of Datavalley, and we should look in the house number 1234."* He showed a map to Fimble in the book. *"This is where you should go".*

"*Are you sure it's still there?*" said Fimble, not wanting to waste a single more step on hunting down data chests he could not find.

"*Absolutely. These books are magic. If someone moves the data chest to a new location, the book will know."*

"*Great*" said Fimble, and took off in a sprint. He soon returned happy carrying a data chest.

But not everything was bliss. Another elf named Faelar never came back. His only clue was to go and talk to somebody named Zhang Wei asking for his data chest. As far as we know, Faelar is still walking from door to door talking to people with that name.

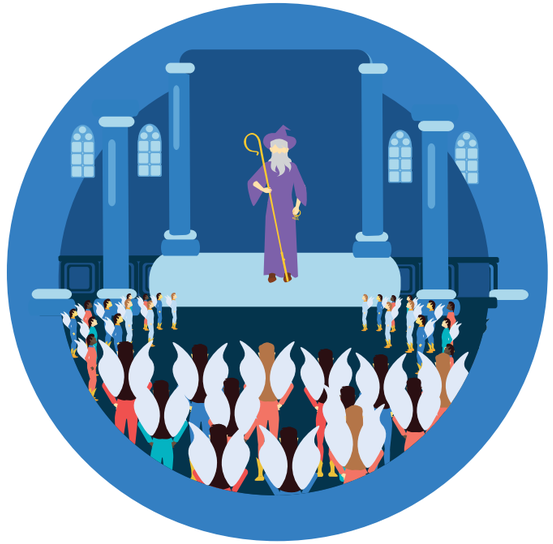
One of the big problems with data concerns the ability to cite your own and other people's data, and keep pointing to its exact location in cyberspace, because the location might change. This can be solved by using persistent identifiers. They work like a big index or registry where you assign a unique key (the identifier) to each data set. If someone tries to follow the identifier (often referred to as "resolving a persistent identifier"), the resolver will point to the correct web address (URL). If the URL changes – i.e. if data is moved –the one who made the key is responsible for providing the new location to the resolver. In this way, you do not end up in blind alleys of "page not found". That is why it is called *persistent*.

A DOI (Document Object Identifier) is an example of such an identifier and looks like this; 10.1234/abba (prefix/suffix). This can be resolved and points you to the URL. Most data repositories can issue and maintain DOIs or other persistent identifiers that you can use. Persistent identifiers usually contain some basic descriptive metadata such as title and author.

Other persistent identifiers are used to prevent ambiguity, e.g. giving a person a number instead of a name. This solves the problem of distinguishing Sam Smith from Sam Smith - yes, there is more than one person with that name! ORCID is an example of a service, where a person is assigned a unique code for further reference, e.g. '0000-0002-1825-0097' that will point to one - and only one - person.

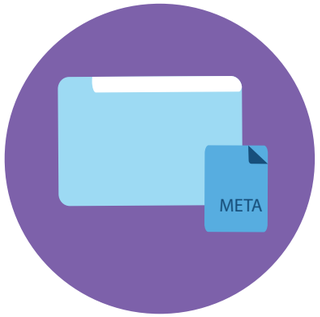
Findable #2:

Data are described with rich metadata



When the first elves returned to the castle, their findings looked far from promising. They gathered around the senior data wizard in the castle hallway. He stepped onto a podium to say a few words: *"You are on the most important mission of all. However, there are a few rules and words of advice that you must keep in mind. When you go looking for chests, you must stick to the paths that you already know from your studies - we cannot find you in the forests, if you walk your own way and get lost. When you find a chest, make sure that you read all about the chest on the label to see if its contents are relevant to us. Read it carefully, as some may not be all that clear in stating what is in the chest. And finally, you must never ever open the chest yourself! Leave it to the data wizards".*

Even though the senior wizard knew that the elves would do their best, he kept all six fingers crossed. The elves were bright, indeed; still he doubted they would locate all the relevant chests out there.



When humans and machines look for data, metadata are often the first point of contact, as they are usually indexed in search engines etc. It is often the metadata that determine whether the data set they describe is perceived as relevant or not for a given usage scenario.

If you asked a human being, the same query words they would use to find a data set should be available in the metadata. This is metadata about the context and/or prerequisites for the data set, quality issues etc., as well as a number of discipline-specific data, e.g. sample size, equipment etc. This also includes details about the data set that may not be important to you, but could somehow be used to make your data findable outside your own discipline. So, try to think outside the box when adding metadata to your data.

Findable #3:

Metadata clearly and explicitly include the identifier of the data they describe

Some elves came back without data chests. Instead, they brought pretty little leaves with engraved texts describing the contents of the chests. One of the elves who came back was Faruty, and she was very proud of the golden leaf she had found.

"Look", she said to the data wizard named Wiux. *"I found this, and it describes the most fascinating chest I have ever heard of. Tell me where to go, and I will fetch it."*

"Let me see", said Wiux turning the leaf over and over. *"Strange …. There are no details about where to find the chest. Are you sure there was no chest next to the leaf?"*

"Yup, sure, sure", said Faruty. *"I found this one right at the outskirts of the enchanted forest, and I promise that there was not a single chest in sight. And I looked in all directions – and dimensions."*

"Then I have no clue where to look", Wiux said. *"My best suggestion would be to go and ask the frightful Lord of Uguly, if this chest could be hidden somewhere in his toxic swamp."*

"Ha ha ha, you are so funny" said Faruty, rolling her one eye. But the smile soon vanished as she looked at Wiux, whose face did not show the faintest sign of amusement..

"Off you go", said Wiux. *"We really need this one"*, sending Faruty on a quest more dangerous than fighting a Syzx dragon.

Metadata and data are two separate things and should be treated as such. They can have a life of their own. An example is metadata that are harvested for big indexes that do not hold or index the contents of the data files. Just like article metadata that are fetched for indexing in commercial and non-commercial search engines, without them looking at the content of the article itself. If the metadata do not include reference to the data they describe, it is doubtful - especially for a machine - that the data described by the metadata will ever be found.

Findable #4:

(Meta)data are registered or indexed in a searchable resource

In their search for data chests, the elves had been reading all sorts of literature to get indications of where to find them. However, the elf Fusky was really fed up looking for data chests, so he sat down under the apple tree in the castle's courtyard for a short break. He noticed the elf Fyrti walking through the yard carrying a large wooden tree chest.

"What's that Fyrti?" he asked.

"It is gold", Fyrti replied with a big smile on his face. "Or at least it looks like somebody tried to make gold before we did."

"Where did you get it from, and how did you find it?" Fusky asked impatiently to know where to look next.

"This … Well I … found it in the forest. It was just lying there on the ground. I'm sure we don't have it in the books," Fyrti said a bit bewildered. "That's ok, isn't it?"

"Sure. It's great", said Fusky. *"But this means that we have to go and look for chests in every corner of every kingdom. It's going to take decades!"*

"I know", said Fyrti, trying to keep the smile on his face. *"But I have just heard rumours that Wildy the data wizard is talking to the people in the library tower. They might have found a magic chest map."*

"Exciting" said Fusky, almost forgetting that he was sick and tired of looking for data chests. *"With all chests mapped?" crossing three fingers.*

"Maybe", Fyrti replied. *"Who knows …?"*

Traversing the entire internet for (research) data sets is neither feasible nor doable. And it leaves too much room for serendipity, which can be nice in some cases, but not desirable when making structured searches for data sets. Making research data available on project websites etc. usually adds to the risk of only being found by coincidence. Repositories are websites – most often – and represent a common way for building structured indexes of metadata and data sets that are uploaded to the repository. The indexes often adhere to a specific way of describing the data using a common standard. This will allow both repository and other search engines to harvest and index these registries, often aggregating them to larger indexes that eventually can be cross-searched.

Repositories come in many shapes and forms. Some are generic repositories that will take almost any data set, while others are targeted towards specific disciplines or research data types. Repositories are usually owned and operated by institutions, research communities, or private companies. The question of where exactly to deposit your data is a matter of determining the best repository for your specific data set, thereby maximizing its findability and potential. This is often evaluated on a case-to-case basis.

Accessible #1:

(Meta)data are retrievable by their identifier using a standardised communication protocol

Alok was the saddest elf in the whole kingdom of Datamania. Or so he thought. A wizard found him by the Magellan fountain on his way back to the castle.

"What's up?" the wizard asked in a perky way.

"I was sent off to the kingdom of Dovia to read all the descriptions of their data chests and to see which ones we could buy. Alas, I was not able to read a single one. I'm horrible at this!" Alok cried out.

"Don't be so hard on yourself", the wizard replied in a sweet and soft voice, regretting his perky attitude. *"Tell me exactly what happened."*

"Every time I grabbed a chest to read what was inside, all the letters on the label changed position. I tried to follow them, but they just kept changing. I'm supposed to be able to read them, aren't I?" Alok muttered.

"Oh, you have come across some old Dovian data chests written in chunky Dovianic. It is complete nonsense to us – and to them as well, I might add. Nobody can read chunky Dovianic, and it is a shame that we cannot even read the chests' descriptions. The last creature that was able to read it was a toad, but he passed away some years ago."

"This is so complicated," Alok sighed. *"I should go and sleep for a couple of days. My eyes are really tired."*

"Good idea", said the wizard. *"And keep up the good work"*.

Once someone has found either your metadata or the data themselves, they - or their machine - should be able to access the (meta)data using standardised mechanisms. This principle states that access should be provided through a standardised protocol. Most often, these are protocols we know from the internet – e.g. http(s) or FTP. This is usually the case, when data are deposited in a trusted repository. However, there might be cases where you will need additional mechanisms such as contract information or similar before someone can access your data. This is perfectly in line with the FAIR principles, if you clearly account for this in the metadata. This may be in the form of contact options that are broadly accepted and easy to use. Examples of this are telephone numbers and email addresses.

Accessible #1.1:

The protocol is open, free and universally implementable

The elf Agon came back to the castle, looking like she had just seen a dragon with four heads.

"I need a cat from the kingdom of Stiodor, a rope, a herring and a bulb horn", she said.

"Excuse me," the data wizard replied. "Why do you need these things? Are you going to combat a Codun dragon?"

"No", the elf replied. *"Heavens no! I followed an untrodden path to get hold of a specific chest. Suddenly, a Gryrvos goblin was standing right there in front of the chest commanding me to swing a cat in a rope, while eating a herring and stepping on a bulb horn. Otherwise would not be allowed to pass the bridge to get the chest".*

"Hmmm", said the wizard. *"I really don't think we have time for that. I can get you a rope, a herring and a bulb horn, but the cat …. Does it have to be from Stiodor? Besides from being very rare, they are extremely expensive."*

"I know", said Agon. *"But without it, the goblin won't let me pass the bridge. No Stiodor cat, no passing."*

"I don't think we can get a cat from Stiodor. And certainly not, if we tell them what we intend to do with it", replied the wizard somewhat vexed. *"You must go and look for another chest, and hopefully we will be able to do without this one.*

A protocol is the technical term for the standard of how data are transported via a network, e.g. the internet. TCP/IP is an example of an open, free and universally implementable protocol that is used as protocol for most of the internet. This means that anyone can use it without having to pay usage fees. If you choose a protocol with restrictions on usage, you might prevent other people from accessing to your (meta)data, thus making it hard – or even impossible – to use the data you have published. If you use a trusted repository for data publication, the repository will make sure that you are in line with this principle.

Accessible #1.2:

The protocol allows for authentication and authorisation when required

The elf Albon was out on one of the most dangerous quests – to get a data chest guarded by the dragon Guardo. Nobody really knew the temper of the dragon, but Albon was not afraid. He was bringing gold and sacrificial gifts to Guardo to get access to the data chest. Furthermore, Albon carried a gold leafed certificate, signed by Prince Fairhair, to prove that the prince had indeed sent him. And Myrtimar, the troll owning the data, had cast a magic spell on the certificate

"Welcome", puffed the dragon, as Albon approached him. *"I'm guarding this data chest, and only those, who can prove themselves worthy, will be allowed to carry the chest from here."*

"Well ...", said Albon *"I carry this signed and spell-cast certificate, and I also know the secret passphrase."*

"Give me the certificate, and tell me the phrase", said Guardo, looking in five different directions simultaneously.

"Here", said Albon handing the certificate to the dragon. *"And the secret passphrase is; Gold to water is so much smarter"*.

The dragon turned around and looked at the small shed behind him.

"Give me the gifts. It's in there for you to take. But remember, only Prince Fairhair can open the chest. If you try, you will burn," said Guardo in a menacing voice, and the door to the shed opened revealing a shining chest of valuable data.

One of the most widely believed myths on FAIR data is that FAIR data must be available as open data, meaning that they should be free for everyone to download. That is not the case.

This principle states that if you place data behind some sort of digital wall, being a paywall or a simple approval system for access, the system must allow for some type of authentication and authorisation. This holds for both humans and machines.

Authentication is all about telling a system who you - or your system - are. Authorisation, on the other hand, is the process where the system is evaluating, if you are allowed to access a given resource.

There are many good reasons to place data behind authentication and authorisation mechanisms. But remember, if you have manual processes involved in evaluating who should have access to the data, the repository or storage system must be able to contact you and seek approval. Other systems will let users or systems register themselves and then provide access to the data, while maintaining a logbook of who has been given access to the data.

Accessible #2:

Metadata should be accessble even when the data are no longer available

The elf Ahlert came back to the wizard Uku with tears in her eyes.

"What's wrong?" inquired the wizard and studied the crying elf.

"It was there. I know it was there, but I couldn't find it", sobbed Ahlert.

"What?" asked the wizard. *"What was where? I don't understand."*

"I had a perfect clue for a data chest. I was so keen on finding it. But all I found were some empty chests with tiny chunks of gold and a kind of mixture lying about. But there was absolutely nothing that looked like data", wondered Ahlert, shaking her head. *"So close, yet so far. It was mentioned in a book ... "*.

"Terribly frustrating!" exclaimed the wizard. *"Did you look for clues out there?"*

"Did I look for clues?" Ahlert replied indignantly. *"I have spent hours looking around, even trying to talk to people from Ogohu Isle where the chest was supposed to be. But nothing. Nothing!"*

"You have to get back in the saddle," ordered the wizard.

"I have a horse?" Ahlert asked bewildered. *"Now I'm really confused."*

"Let me explain something to you," the wizard said to Ahlert, placing a hand on her shoulder and accompanying her out of the basement. *"Horses aren't always in flesh and blood ..."*

There are plenty of good reasons why data disappear from repositories and similar places. They may be withdrawn due to cost of having them online, if no one is left to maintain access to the data, or for other reasons. However, if the metadata about the data set disappear as well, it will leave humans and machines in an "unfulfilled" state, when they try to retrieve the data, e.g. when resolving a persistent identifier, or following a link to the entry of a repository. Therefore, make sure to leave metadata telling that "yes, the data were here, but they are no longer available". Leaving metadata may also offer information on the context, the authors and the institution, where the data were created, for those looking for further details.

Interoperable #1:

(Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation

It did not take long, before all rooms in the basement of the castle were swamped with huge piles of data chests from kingdoms close to Datamania and empires far far away. The wizards were struggling to keep up with all the chests that the elves brought back to the castle. However, the sheer mass of data chests was not the only source of frustration. It quickly became clear that not all chests had something to do with turning water into gold although both *"water"* and *"gold"* were on the label. The wizard Igly shook his head in despair. He was concerned, and they had not even begun to look for chests with the labels "gold" and "H_2O".

"What's wrong?" asked Ilo, a small elf with a squeaky voice, gently padding the wizard on the head.

"Look around you", sighed Igly. *"Look at all these chests. It is good that you bring them here, but many of them are not relevant at all. And I fear that some of the relevant ones are not found".*

"I see", replied Ilo, although she didn't really understand the wizard's concern. *"What's that sound? Is it... music?"*

"Oh, that sound", the wizard answered looking almost betrayed. *"That is music. Someone came back with a data chest labelled Gold, but it turned out to be music by the lizard Björn Wolfsson".*

"Sweeeeeet", Ilo screamed so loud that it could be heard all the way through the castle corridors. *"Let's have some fun and dance for three days and nights. Isn't this great?"*

"Not really, you see...," Igly mumbled, but Ilo was long gone

Humans, and especially machines, can have a hard time interpreting data. Words are ambiguous, and the multitude of spoken and written languages add further to the complexity. Problems range from not being able to interpret the value of a cell due to missing information on the metrics used, to more complex situations where you will have to look for many different terms describing the same object, or stumble across words with different semantic meaning across various disciplines. The same holds for place names and the like.

The FAIR principles address this issue by recommending the use of shared data standards for representing data, and the use of vocabularies and ontologies to represent values and mark-up data. Vocabularies and ontologies are often defined within the research communities and are an unambiguous way of adding semantic meaning to your data. A simple example is to rely on a flower ontology to classify flowers, instead of writing their names in plain text.

Working with data in this way can make your data more useful and discoverable. However, you should be aware that this type of work often has an impact on your methods and the software you use.

Interoperable #2:

(Meta)data use vocabularies that follow the FAIR principles

The elf Imka had found a data chest that the data wizard Dorky was about to open. Imka was so curious that he was allowed to stay and watch as Dorky opened the chest.

"Look", exclaimed Imka, as they opened the chest. *"It holds Carmix bubbles. Isn't that amazing? These can certainly help in turning water into gold, right?"*

"Yes, absolutely", said Dorky. *"But take a closer look, they have markings on them. These are so-called Polymixic markings. Only very few wizards – if any – know how to interpret these. They indicate how each of the Carmix bubbles are associated to the others, and how to place them in the right order."*

"Polymixic markings … I have never heard of them", said Imka scratching his head.

"I only know them from a tale, and I actually did not believe in their existence," said Dorky gloomily. *"Once, a wizard named Yrky worked at the castle, and he apparently knew a wizard who could sort these Polymixic markings. But he died only 523 years old."*

"But he must have left something that can help us", Imka cried out in a passionate voice. But the expression on Dorky's face made him fall silent.

"I'm sorry, little fellow", Dorky muttered apologetically. *"Without the right understanding of the logic of Polymixic markings, we can't really do anything. But they taste deliciously"*, he said, as he sank his teeth into one of them, and the bubble emitted a little fizz, as he began to chew.

A vocabulary is only good, if it is accessible and allows for the right interpretation of the data. This principle highlights the importance of using vocabularies that are common to the community and well documented, and can be referred to using persistent identifiers. Usually, you will find this type of vocabulary, taxonomy etc. within your research discipline or maybe in other disciplines, where these are developed. Evaluating a vocabulary often includes looking for its creator and checking whether it is still maintained and updated. These can be complex vocabularies; or simple mark-ups like ISO standard strings for representing countries in a data set. E.g. Denmark is DNK in ISO 3166-1 alpha-3. In terms of interoperability, this is far better than writing 'Danmark', 'Denmark', 'Dänemark, 'Dinamarca' etc. in your (meta)data.

Interoperable #3:

(Meta)data include qualified references to other (meta)data

The elf Inandu returned all excited to the castle. He had found a data chest with a correct label with the correct location of where he later found the chest. This chest contained a magic spell of how to turn water into gold.

The überwizard Ikloton opened the data chest and started reading through the spell.

"This is good", he rejoiced. *"Really, really good. Inandu, I think this is the end of our quest."*

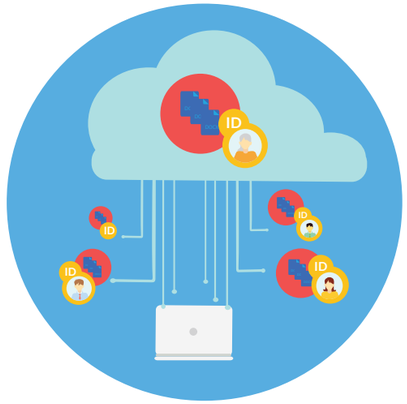
"Should I call upon the prince?" the elf asked, blushing with pride that he had brought this chest to the castle.

"Yes … But wait!" the überwizard grunted, first turning white, then red with anger: *"In step 4 of the spell, the wizard authors refer to a procedure they have used to conjure the key ingredient used in this spell - violet lizards with green dots. What is this procedure called? And how can it be found?"* The spell gave no indication of this, neither did the data chest or its label. Thus, even though the data chest contained the right spell, it was useless to the data wizards at the castle.

"I'm … so … sorry", mumbled the elf, his eyes brimming with tears. He felt like crying for days, which he actually could.

"Me too", said Ikloton. *"So close, and yet we are nowhere"*.

It is important to be able to trace the connections between your data set and data sets that are related to it. This can be done by linking to other data sets that are not included in your work. It can also be done through connections that show how your data set is derived from a previous version or e.g. is processed data based on some raw data. Either way, it is important to maintain these connections by referencing between data sets. If your data set relies on other's data - or your own - this is also an appropriate method to ensure that proper credit is given to the people who created the data that your data is based upon.



Reusable #1:

(Meta)data are richly described with a plurality of accurate and relevant attributes

The elves were bringing back data chests to the wizards like crazy, leaving both the elves and the wizards in an almost frantic state of mind. The wizards were struggling to figure out the contents of the chests. One of the elves that came back with a chest was Roscoe. Proud as most elves, he brought a nicely wrapped chest to the wizard, and one he had paid for, too.

"Hmm", the data wizard muttered. *"What's this? Hopefully better than the last one?"*

"Well …" said Roscoe. *"It was hard for me to find out, if it was right for me to bring it back here. I know we were supposed to bring back only the ones that succeeded in turning water into gold. But this doesn't say what it has been used for".*

"Geez", said the wizard a little distressed. *"I have to look into it to figure out if it's relevant. But this will take time, and I do not have time."*

"Sorry", apologized Roscoe looking down.

"We don't have room for more irrelevant chests in here", exclaimed the wizard. "You'll have to focus on the ones that we know will transform water. And don't pay for anything that we are not sure of."

"What …" said Roscoe doubtfully. *"Leave … chests … behind …?"*

"Yes", said the wizard in a firm voice. *"I know it is hard, but we have to make sure that we only bring in relevant chests."*

"Alright", said Roscoe. *"I think I'll catch a fishing rod on my way out and see if I can find a trustworthy data lake".*

Labelling your data with relevant attributes - most often in the form of metadata – does not only help discovering your data. It also helps humans and machines to understand the context of your data. This can be in the form of purpose and processing statements, equipment used, software versions etc. Imagine finding your own data. Now think of the contextual information that would benefit you in determining whether the data is relevant to your specific needs - and whether you would be able to understand how the data were created. Be generous when adding attributes to your data. What might not be relevant to you might be the part of filtering and querying for data for other people - or machines.

Speaking of machines, your best choice would be to use e.g. controlled vocabularies, persistent identifiers or similar to make the contextual description unambiguous. Often repositories targeted towards specific disciplines, communities, or data types will have the most optimal support for both assigning, maintaining and querying using domain specific metadata.

Reusable #1.1:

(Meta)data are released with a clear and accessible data usage license

The data chests brought back by the elves came in all different sizes, shapes and packaging. Some were even locked in chains with a padlock. Luckily, the elves were able to buy a key for them. Others were nicely wrapped with curled ribbons and small tags, saying things like 'For you to keep', and even marked with small drawings to help the elves where they could not read the local tongue. One elf named Ruby came back with a trembling data chest.

"What's wrong with this one?" Ruby asked the wizard. *"It is wrapped like the others, but there is no tag on it. Have I done something wrong?"*

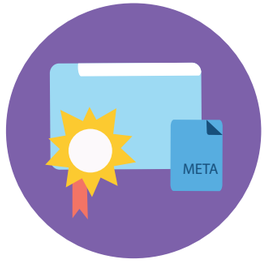
"Oh, let me see", said the wizard. *"It looks exactly like what we need, but we have to throw it in the moat. We do not know if it is bewitched. And we can't use it without knowing whether we are allowed to use it or not."*

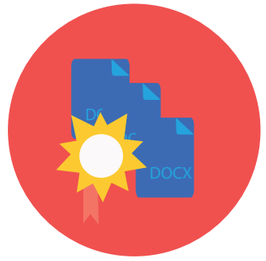
"What!" exclaimed Ruby in disbelief. *"Can't you do something? If this is what we need, why can't we keep it?"*

"I know, I know", said the wizard. *"Although it looks like a gift, we can't keep it, because it doesn't carry the gift tag. But this is not your fault Ruby"*, the wizard said padding Ruby on the head.

Licensing data and metadata is an important aspect of the FAIR principles; both if you retain some or all rights, or if you set your data free as completely open data. In a FAIR context, a license is a standardised machine-readable statement that tells the end user exactly how he or she can use the data, and under which conditions.

You can apply many different licenses to data. One of the most common is the Creative Commons license suite, where you explicitly state how and if you are to be cited when your data are reused, along with possible re-sharing options of derived works etc. Typically, you choose a license when depositing data in a repository. You do not have to apply the same license to all your data, and stating full copyright is a license in itself. The worst thing you can do is not to apply any license at all. Because then your data is most likely protected by copyright as default, even though the data are publicly accessible.





If you use other people's data, you should always identify the conditions of how and when you can use the data. This may affect the way you can work with their data.

Reusable #1.2:

(Meta)data are associated with detailed provenance

The elf Rherek hurried home to the castle as if running on a gust of magic wind.

"I got it, I got it, I got it", he shouted almost out of breath.

"Got what...?" said Jimko, the data wizard. *"Take a deep breath, relax and tell me."*

"I ... got ... the right recipe for turning water into gold", he whispered. *"We need a giant with three heads. And this chest shows how to conjure up such a giant."*

"Marvellous" said Jimko. *"Let's look at the chest immediately."*

They investigated the chest. The content looked like playing cards. However, they were all mixed up.

"Now ..." said Jimko. *"This looks perfect. It is in line with what we believe to be the right steps. However, it is all mixed up, so we have no clue in which order to take the steps. And it gets even worse ..."*

"How is that?" asked Rherek. *"Can't we just try it in different sequences? We must succeed eventually."*

"Or blow up the castle. If you do anything wrong with the giant, it will turn more evil than a Buffingor witch on a bad day!"

"But, but, but ... can't we ask those who created the cards?" Rherek asked warily.

"We can... But the chest contains no trace of who created the cards. It will be a stroke of luck. But try and go back and look for traces of evidence of who created the chest", said Jimko despairingly.

Rherek ran off. But neither he nor Jimko were at all convinced that further investigation would make it possible to understand the cards.

Much of the value in data is the ability for a machine or a human to judge the origin of the data. Thereby often evaluating whether the data is reusable in a new context. This includes the ability to know how the data was created, by whom it was created, and with which type of equipment? Also, has the data been processed, or is it raw data? If it was processed, how was the workflow? And so on, and so on. This is quite similar to a section on method in a paper or article, and you can refer to this type of documentation from your data set. However, keep in mind that this might not be readable for a machine.

Remember to include provenance of who you are, and how you would like the data set to be cited/credited if used elsewhere.

The easiest way to get started, is to try and think of yourself as a re-user of your own data. But before you do so, you must clear your head of all knowledge related to the data set. What details would you need to evaluate and trust a given data set? If this is hard to imagine, try finding other people's data sets, and see if you think they have enough provenance.

Reusable #1.3:

(Meta)data meet domain-relevant community standards

Rebzuss was the last elf returning with a data chest. She was glowing with expectation and pride, because she had found a data chest containing a spell for turning water into gold. The data wizard Fixeor Datahin looked at the chest label:

"Well done, Rebzuss. This is exactly what we are looking for". He opened the chest and started frowning:

"Hmm ... I can see that this is the right spell, but I can't quite understand it. It looks like the spell is written by a data wizard trained at the data lab at Oxwart University. I recognize the peculiar use of logical symbols and the Oxwartish way of data handling in procedure 1 and 5. It will take us years to translate this spell into Scruby."

Rebzuss looked sad - that was not the reaction she had hoped for. Suddenly, her face lit up and she said:

"Overwizard Fixeor, why don't we call upon the witch Lux Datastorm. Before working at the castle in Datamania, she studied at the data lab at Oxwart. It might be easy for her to translate the spell for us."

The overwizard immediately summoned Lux Datastorm and showed her the data chest and its contents. She laughed: *"I can see why you have difficulty understanding the spell. This is described just the way they do spells at Oxwart. Give me an afternoon and I will translate it into Scruby and our Datamanish procedures."* Lux took the data chest and disappeared into her chamber.

Working with data sets from a variety of resources is much simpler, if everybody agrees to a certain standard way of organising and describing the data. That is why many disciplines have created metadata standards for describing data, and created lists of recommended file formats etc. Keeping in line with these standards will lead new data out into the ecosystem of data that is easy and suitable for others to reuse. Therefore, you should always try to be on the lookout for standards within your community and try to adhere to these. However, not everything can be standardized, of course, and many research disciplines are breaking new ground where there are currently no standards – and then you will turn to more generic standards, or begin inventing new ones.

Notice that standard in this sense is not a quality measurement indicating a level of high or low quality of the (meta)data. It should always be judged by the people who are re-using the data in their specific contexts.

Ending

The following day Lux Datastorm returned with the recipe on how to turn water into gold. All the elves, data wizards, and witches were present when they concluded that the recipe would actually work. Everyone applauded when the recipe was properly stored in a glass chest and brought before Prince Fairhair.

He immediately took his white horse and rode off to the neighbouring kingdom. Three days later, he returned with My Fair Lady. They were married at a great ceremony in the library, and the party was held in the archives - some of the most beloved places in Datamania.

Of course, the quest was documented according to the FAIR principles. In this way, no one would ever have to go through the same troubles again, and they could all live happily ever after.