

## L1C13 - Bringing it all together

0:01 so in the last several concepts we  
0:02 walked through a series of tools that  
0:04 combined together to build the remote  
0:06 data science ecosystem  
0:08 and where we left off was this this  
0:10 vision for how in theory these tools at  
0:13 a low level  
0:15 can make it possible for a data  
0:16 scientist and **a whole fleet of data**  
0:19 owners  
0:20 to coordinate on an experiment where the  
0:23 data owners don't have to actually  
0:25 really babysit the experiment too much  
0:27 as long as they're okay with that  
0:29 category of experiment happening and can  
0:31 allocate a certain degree of privacy  
0:32 budget to the data scientist  
0:34 and the data scientist can perform this  
0:36 **whole interaction** while  
0:38 sort of it feels like the data is on the  
0:40 data scientist's own machine meaning  
0:43 that the the user experience for the  
0:45 data scientist changes as little as  
0:48 possible relative to  
0:50 how the data scientist would have worked  
0:52 if they were requiring a copy of the  
0:54 data  
0:55 but this still leaves  
0:57 a little bit  
0:59 lacking in terms of  
1:00 okay that's how the technology works but  
1:02 how does the technology actually  
1:04 interface  
1:05 with the in the the institutions who are  
1:07 doing the approving right in the sense  
1:10 of of there's still a data owner who has  
1:13 a certain set of wishes about how they  
1:14 want to protect you know the data in  
1:16 their own institution and and probably  
1:18 not participate in data science projects  
1:20 that might have for example ethical  
1:22 problems or competitive problems or  
1:23 whatever else it might be and of course  
1:26 while not sharing the data outright  
1:28 makes that decision a lot easier it  
1:29 doesn't necessarily limit the entire  
1:31 need to to facilitate that decision  
1:33 there's also some other sort of just  
1:35 just more institutional level  
1:37 experiences that are slightly different  
1:38 so i want to introduce a couple of  
1:40 important concepts just before we finish  
1:42 out the lesson they're going to be  
1:44 really important to how you you learn  
1:45 about this throughout the rest of the  
1:47 course because really unpacking exactly  
1:48 how these different institutions play  
1:50 together i mean that's that's a big part  
1:51 of what this course is about so we're  
1:52 not going to cover all of in this lesson

ここまでのいくつかの概念では、  
一連のツールについて説明しました。  
組み合わせてリモコンを構築する  
データサイエンスエコシステム  
そして中断したところはこれでした  
これらのツールが理論的にどのように機能するかについてのビジョン  
低いレベル  
データを可能にすることができます  
科学者と**データ群全体**  
所有者  
実験に関して調整するため、  
データ所有者は実際にはそうする必要はありません  
本当に実験の子守りが多すぎる  
彼らがそれでいいなら  
行われる実験のカテゴリ  
ある程度のプライバシーを割り当てる  
データサイエンティストに予算を与える  
データサイエンティストはこれを実行できます  
**インタラクション全体**  
データがそこにあるような気がします  
データサイエンティスト自身のマシンの意味  
のユーザーエクスペリエンス  
データサイエンティストの変更は最小限で済みます  
相対的に可能  
データサイエンティストはどのように働いていたろうか  
のコピーが必要な場合  
データ  
しかし、これはまだ残っています  
若干  
～の点で欠けている  
まあ、それがテクノロジーの仕組みですが、  
実際にその技術はどうなっているのか  
インターフェース  
いる機関の中で  
ある意味で正しい承認を行っている  
のデータ所有者がまだいます  
彼らがどのように行動するかについての特定の願望  
保護したいのですが、次のデータを知っています  
彼ら自身の機関、そしておそらく  
データサイエンスプロジェクトには参加しない  
それは例えば倫理的なものかもしれない  
問題や競争の問題、あるいは  
それが何であれ、もちろん  
**データを完全に共有するわけではありませんが、**  
その決断がずっと簡単になります  
**必ずしも全体を制限するものではありません**  
その決定を促進する必要がある  
他の種類のものもある  
より組織的なレベルで  
少し違う経験  
そこでいくつか紹介したいと思います  
終了直前の重要なコンセプト  
彼らがこれから受けるであろう教訓を  
学び方にとって非常に重要  
これについては残りの部分で説明します  
もちろん、本当に正確に開梱するので、  
これらのさまざまな機関がどのように役割を果たすか  
一緒に言うと、それが大きな部分だと思います  
このコースの内容については  
このレッスンではすべてをカバーするわけではありません

1:54 but there are few key concepts that that  
1:55 we will we will kick off with here so  
1:57 the first is  
1:59 the difference between a network and a  
2:01 domain node so  
2:03 um  
2:04 when i've been saying take your data  
2:06 as a data owner you know at a hospital  
2:08 or something like that put it into a  
2:09 server  
2:10 that is owned and operated by the  
2:12 hospital and advertise it to sort of the  
2:14 internet i've been referring to what's  
2:16 called a domain node meaning that this  
2:19 is their domain their data continues to  
2:21 live  
2:22 inside of this particular server right  
2:25 the the analogy this is sort of the  
2:26 apache web server for private data right  
2:29 and it's got a web ui and allows for a  
2:32 certain degree of configuration all  
2:33 these types of things and and then  
2:36 there's but there is a second type of  
2:37 node because like  
2:38 there are certain things that domains  
2:39 are not well suited for they are well  
2:40 suited for protecting data and  
2:42 configuration all this type of stuff but  
2:44 there's a second type of node which is  
2:45 called a network node and a network node  
2:48 facilitates services that are a bit more  
2:50 like you know sort of search like search  
2:52 and discovery right um if i'm a data  
2:55 scientist and i need to be able to find  
2:57 diabetes data if there's 10 000 domain  
3:00 nodes out there i don't necessarily want  
3:02 to go to every single domain and be like  
3:03 do you have diabetes data do you have  
3:05 diabetes do you have diabetes do you  
3:07 have diabetes data it's just like kind  
3:08 of a painful um a painful process  
3:11 secondarily um  
3:14 network nodes are going to facilitate a  
3:17 second really important  
3:19 sort of sort of role and this will  
3:20 necessarily happen immediately but in  
3:22 terms of their long-term role of  
3:24 services to to the network and that role  
3:26 is is one of delegated permissions so  
3:28 let's say  
3:30 i'm going to walk through a little bit  
3:31 of how  
3:32 academic institutions do data approvals  
3:34 now and you'll start to see the analogy  
3:36 um  
3:37 so  
3:38 if i wanted to work with the data that's  
3:40 at another university um what i'll end  
3:42 up doing is i'll sign  
3:44 a data sharing agreement and i'll also  
3:46 have to submit my project for irb  
3:48 approval so an irb board is usually  
3:50 typically a collection of of professors  
3:53 at that university who review research

しかし、それを実現する重要な概念はほとんどありません。  
ここからキックオフしますので  
一つ目は  
ネットワークとの違い  
ドメインノードなので  
えーと  
私があなたのデータを持ってって言ったとき  
病院で知っているデータ所有者として  
またはそのようなものを  
サーバ  
によって所有および運営されています。  
病院を宣伝し、  
インターネット、私が言及してきたものは何ですか  
ドメインノードと呼ばれる、これはこれを意味します  
彼らのドメインは彼らのデータであり続けます  
ライブ  
この特定のサーバー権限内で  
これは一種のアナロジーです  
プライベートデータ権のためのApache Webサーバー  
ウェブ UI を備えており、  
ある程度の構成はすべて  
こういった種類のこと、そしてそれから  
それはありますが、2番目のタイプもあります  
ノードのようなので  
ドメインには特定のものがあります  
あまり適していない  
データの保護に適しており、  
この種の設定はすべて行いますが、  
2 番目のタイプのノードがあります。  
ネットワークノードおよびネットワークノードと呼ばれる  
もう少し便利なサービスを提供します  
ご存知のように、検索のようなもの  
そして、もし私がデータなら、それを発見してください  
科学者と私は見つけることができる必要があります  
ドメインが 10,000 個ある場合の糖尿病データ  
そこにあるノードは必ずしも必要ではありません  
あらゆる単一ドメインに行き、次のようになります  
糖尿病のデータはありますか  
糖尿病 あなたは糖尿病を患っていますか  
糖尿病のデータがある、それはまさに親切だ  
痛みを伴う、ええと、痛みを伴う過程の  
二次的には、ええと  
ネットワークノードは、  
2番目に本当に重要  
ある種の役割であり、これは  
必然的にすぐに起こりますが、  
長期的な役割の観点から、  
ネットワークへのサービスとその役割  
これは委任されたアクセス許可の 1 つであるため、  
まあ言ってみれば  
少し歩いてみます  
どのように  
学術機関がデータ承認を行う  
そうすれば、類似点がわかるでしょう  
えーと  
それで  
データを操作したい場合は、  
別の大学で、えー、何を終わらせませうか  
やるべきことはサインします  
**データ共有契約**を結びます、そして私もそうします  
IRBにプロジェクトを提出しなければなりません  
承認が必要なため、通常は IRB 委員会が行われます。  
通常は教授の集まり  
研究をレビューするあの大学で

3:55 projects to decide whether it's it  
3:58 surpasses ethical reviews it's something  
4:00 that they they think is a good use of  
4:01 their resources and and across lots of  
4:04 you like all the universities in a  
4:05 particular nation or jurisdiction  
4:08 there'll be standards by which ethical  
4:09 reviews happen meaning the ethical  
4:11 review at you know for oxford university  
4:13 here should be the same as the  
4:15 university of iowa in the united states  
4:17 in terms of the principles that they use  
4:19 for for ethical review  
4:21 and so this brings us to  
4:22 remote data science  
4:25 even though the technology technically  
4:28 exists  
4:29 for me to be able to do data science  
4:31 across a thousand universities right or  
4:33 do data science across data hosted by a  
4:36 thousand universities  
4:37 the  
4:38 logistical red tape  
4:40 was  
4:41 technically require me to go for irb  
4:44 review  
4:45 at a thousand different universities  
4:48 and that's  
4:50 wildly redundant in the sense that  
4:53 i'm gonna get the same review a thousand  
4:55 times over and look at the same document  
4:57 with the same procedures and and  
4:59 in theory make the same decision  
5:02 um  
5:03 which involves you know 4 000 people or  
5:05 5 000 people all reviewing my proposal  
5:08 when  
5:09 they're all fundamentally making the  
5:10 same decision and so the idea here is  
5:12 that the the secondary service optional  
5:14 service that that network nodes will  
5:16 provide to to networks of domains  
5:20 is is one of delegated irb approval in  
5:23 the sense that  
5:25 a um this is the strongest analogy here  
5:27 is the data consortium so if you're  
5:29 familiar with the data consortium what  
5:31 will happen is a a group of universities  
5:33 will come together and they'll say you  
5:34 know what we're going to put all of our  
5:35 data into one place right for and and  
5:38 we're going to give that that central  
5:40 actor the ability to decide who's  
5:41 allowed to download copies of this data  
5:43 to do research on it right um so it's  
5:45 got a little in common what we're  
5:47 talking about  
5:48 and a data consortium is really all it's  
5:50 trying to do is cut through the red tape  
5:52 you're still downloading copies of data  
5:53 so it's not quite privacy and  
5:55 technologies like we're we're where  
5:56 we've been talking about for the last uh  
5:58 you know nine or ten concepts but it is

それかどうかを決めるプロジェクト  
倫理審査を超えたそれは何か  
彼らはそれを有効に活用していると考えている  
彼らのリソースと多くの  
あなたはすべての大学が好きです  
特定の国または管轄区域  
倫理的な基準が存在するだろう  
レビューは倫理を意味します  
オックスフォード大学のレビュー  
ここはと同じであるはずで  
アメリカのアイオワ大学  
彼らが使用する原則という観点から見ると  
倫理審査用  
これにより、  
リモートデータサイエンス  
たとえ技術的にはその技術であっても  
存在します  
データサイエンスができるようになるために  
1,000 の大学にわたって正しく、または  
によって**ホストされているデータ全体**でデータ サイエンスを行う  
**千の大学**  
の  
物流のお役所仕事  
だった  
技術的には IRB を受ける必要があります  
レビュー  
千の異なる大学で  
そしてそれは  
という意味で非常に冗長です  
同じレビューを何千件も受けるだろう  
何度も同じ文書を見る  
同じ手順で、そして  
理論的には同じ決定を下す  
えーっと  
これには 4,000 人の知り合いがいるか、  
5,000 人全員が私の提案を検討しています  
いつ  
彼らはすべて根本的に  
同じ決定なので、ここでのアイデアは次のとおりです  
セカンダリ サービスはオプションであること  
そのネットワークノードが提供するサービス  
ドメインのネットワークに提供する  
は、委任された IRB 承認の 1 つです。  
という感覚  
うーん、これが最も強力な例えです  
は**データ コンソーシアム**です。  
データコンソーシアムに精通しているもの  
起こるのは大学のグループです  
一緒に来て、彼らはあなただと言うだろう  
私たちがすべてを注ぎ込むつもりであることを知っています  
データを 1 か所にまとめて、および  
そこに中心を与えるつもりです  
俳優は誰が誰であるかを決定する能力  
このデータのコピーをダウンロードすることを許可します  
それについて正しく研究するために、ええと、それはそうです  
私たちの性格に少し共通点がありました  
について話す  
データ コンソーシアムこそがすべてです  
しようとしているのは官僚的な手続きを突破することだ  
まだデータのコピーをダウンロード中です  
完全なプライバシーではないので、  
私たちがどこにいるかのようなテクノロジー  
私たちは最後まで話し合ってきました  
あなたは9つか10の概念を知っていますが、それは

6:01 um um it is similar in the sense of it's  
6:04 trying to overcome this logistical  
6:05 hurdle of if you wanted to do data  
6:08 science against data from a bunch of  
6:09 institutions you'd have to go through  
6:10 the red tape of every single one right  
6:12 and so what a network node is  
6:14 facilitating in this particular context  
6:15 is like it's the delegated authority  
6:18 where there's sort of a consortia  
6:19 network node that can play this role of  
6:23 doing irb authority irb  
6:25 decisions on behalf of a large number of  
6:27 universities  
6:28 um  
6:29 but the network node doesn't need to  
6:30 acquire a copy of the data in the  
6:32 process to do that right so so um just  
6:36 want to sort of lay out the way this  
6:38 future ecosystem is going to play so  
6:40 what this would mean is that the data  
6:42 scientist needs to find a network node  
6:43 somewhere that has access that has  
6:45 custodial access over distributed  
6:47 federated data networks for data that's  
6:49 they're relevant to this data scientist  
6:51 to approve their project and they would  
6:52 they would have the ability to to do  
6:55 that  
6:56 so um there will be other services that  
6:58 networks or network nodes will will  
7:00 participate or facilitate to the network  
7:02 so things like providing secure vpns for  
7:04 kind of your more private federated data  
7:06 networks so it might be walled off from  
7:08 from the rest of the internet um and  
7:10 it's likely that that this is going to  
7:12 sort of unfold over time right but the  
7:14 the general principle is that domain  
7:16 nodes are servers for individual data  
7:18 owners at a time  
7:19 network  
7:20 node runners are still there's another  
7:22 type of server on the network but that  
7:24 facilitates sort of services to the  
7:26 network right in that sense so network  
7:27 node will not have data attached to it  
7:29 it will instead help facilitate data  
7:32 scientists finding and working with  
7:33 large groups of domain nodes the other  
7:36 thing i really want to  
7:38 raise here is that that  
7:39 network nodes and domain nodes should be  
7:41 overlapping a domain node should  
7:42 definitely definitely definitely be  
7:44 joining multiple different networks  
7:46 which are probably grouped by theme so  
7:48 you can think there's an irb that's like  
7:49 focusing on breast cancer so it'd be a  
7:51 breast cancer network that would really  
7:52 focus on on doing that maybe it would be  
7:54 by geography you know you have a uk  
7:56 network nodes or ua clusters of network  
7:58 nodes

うーん、意味的には似ていますね  
このロジスティックスを克服しようとしている  
データを作成したい場合のハードル  
多数のデータからの科学  
通過しなければならない機関  
一人一人の官僚的な手続きは正しい  
ネットワークノードとは何ですか  
この特定の状況での促進  
委任された権限のようなものです  
コンソーシアムのようなものがあるところ  
この役割を果たすことができるネットワーク ノード  
irb権限irbを行っています  
多数の人々を代表して決定を下す  
大学  
えーつと  
しかし、ネットワークノードはその必要はありません  
内のデータのコピーを取得します。  
それを正しく行うためのプロセス だから、まあ、ただ  
このようにレイアウトしたい  
将来のエコシステムはそうするだろう  
これが意味するのは、データが  
科学者はネットワークノードを見つける必要がある  
アクセスできる場所  
**分散型**の保管アクセス  
データのための**フェデレーション データ ネットワーク**  
それらはこのデータサイエンティストに関連しています  
彼らのプロジェクトを承認すれば、  
彼らには～する能力があるだろう  
それか  
他にもサービスはあるでしょう  
ネットワークまたはネットワーク ノードは、  
ネットワークに参加または促進する  
安全なVPNを提供するなど  
より**プライベート**な**フェデレーション データ**の一種  
**ネットワークから隔離**されている可能性があります  
インターネットの残りの部分から、ええと、  
おそらくこうなるだろう  
時間の経過とともに展開するものですが、  
一般原則はそのドメインです  
ノードは個々のデータのサーバーです  
一度に所有者  
通信網  
ノードランナーはまだ別のものがあります  
ネットワーク上のサーバーの種類ですが、  
に対するある種のサービスを促進します  
ネットワークはその意味で正しいので、ネットワーク  
ノードにはデータが接続されません  
むしろデータの円滑化に役立ちます  
科学者が発見して協力する  
ドメインノードの大規模なグループ、その他  
私が本当にしたいこと  
ここで上げるのはそれです  
ネットワーク ノードとドメイン ノードは次のようにする必要があります。  
ドメインノードをオーバーラップさせるには、  
絶対に絶対に絶対そうなる  
複数の異なるネットワークに参加する  
おそらくテーマごとにグループ化されているので、  
次のような irb があると考えられます  
乳がんは焦点を当てているので、  
乳がんネットワークは本当に  
そうすることに集中する  
地理的には英国があることがわかります  
ネットワークノードまたはネットワークのuaクラスター  
ノード



7:58 and domain should be subscribing to as  
8:00 many of the network nodes as they can  
8:01 that might have access or might be  
8:03 gateways to data scientists that will be  
8:05 relevant that domain nodes servers if  
8:07 that makes sense so now the domain nodes  
8:09 right they store data as mentioned  
8:11 before they also process data and they  
8:13 also facilitate data governance so so a  
8:15 domain node is where a data owner can  
8:16 put in their preferences you know and  
8:18 their preferences might be you know  
8:20 constraints on how privacy budgets are  
8:22 allocated it might be the domain node  
8:24 decides to to allocate privacy budgets  
8:27 specifically for you know for explicitly  
8:29 for any data scientist working with  
8:30 their data and it's also domain node  
8:32 deciding hey what networks do they want  
8:33 to delegate partial authority to for for  
8:35 different topics or to work with  
8:36 different parts of their data right  
8:38 so there's a lot more to unfold um  
8:41 both  
8:42 in the later lessons of this course and  
8:44 and in the coming years as remote data  
8:46 science ecosystems really build out but  
8:48 um i i want to finish by sort of  
8:49 returning back to the core motivation of  
8:52 what this is really all about right it's  
8:54 all about bringing all the world's  
8:55 information to bear on our most  
8:56 important problems and privacy enhanced  
8:58 technologies and remote data science are  
9:00 really they're just just a  
9:02 the next wave of tools right that are  
9:05 going to reduce the amount of friction  
9:07 increase the amount of safety and  
9:08 security through which we can do that  
9:10 it's gonna be an incredibly exciting  
9:11 time i hope that you see both a personal  
9:13 opportunity for yourself in terms of for  
9:15 your own career as a data scientist or  
9:17 data owner  
9:18 that you'll be going to do  
9:19 groundbreaking research as a result of  
9:21 having access  
9:22 to orders of magnitude more information  
9:24 than other data scientists otherwise  
9:25 would the significant first mover  
9:27 advantage is to you you're being so  
9:29 early  
9:30 in this way by taking this course  
9:32 and that you also see sort of the moral  
9:34 and ethical  
9:35 positives and moral ethical reasons why  
9:37 you helping to push  
9:39 into this direction and push you know  
9:41 the academic communities and government  
9:43 communities and the private sector  
9:44 communities into this direction of  
9:46 adopting these technologies as opposed  
9:47 to creating these centralized data you  
9:50 know databases right that are more

ドメインは次のようにサブスクライブする必要があります  
可能な限り多くのネットワークノード  
アクセスできる可能性がある、またはアクセスできる可能性がある  
データサイエンティストへの登竜門となる  
ドメインノードサーバーに関連する場合  
それは当然のことなので、ドメインノードは  
そうですね、前述したようにデータを保存します  
データを処理する前に、  
データ ガバナンスも促進します。  
ドメインノードはデータ所有者ができる場所です  
あなたが知っている彼らの好みを入力し、  
彼らの好みはご存知かもしれませんが  
プライバシー予算の制約  
割り当てられているのはドメインノードである可能性があります  
プライバシー予算を割り当てることを決定する  
特にあなたにとっては明示的に知っています  
データサイエンティスト向け  
彼らのデータ、そしてそれはドメインノードでもあります  
どのようなネットワークが必要かを決定する  
部分的な権限を〜に委任する  
さまざまなトピックや作業に使用する  
データ権利のさまざまな部分  
まだまだ展開すべきことがたくさんあります  
両方  
このコースの後半のレッスンでは、  
そして数年後にはリモートデータとして  
科学のエコシステムは実際に構築されていますが、  
うーん、なんとか終わらせたいです  
〜の核となる動機に立ち返る  
これは本当に何ですか、そうですね  
世界中のあらゆるものをもたらすことがすべて  
私たちにとって最も重要な情報  
重要な問題とプライバシーの強化  
テクノロジーとリモートデータサイエンスは、  
本当に彼らはただの  
まさに次のツールの波です  
摩擦の量を減らすつもりです  
安全性の量を増やし、  
それを可能にするセキュリティ  
信じられないほどエキサイティングなものになるだろう  
両方の個人的なものを見ていただければ幸いです  
〜のための自分自身のための機会  
データサイエンティストとしてのあなた自身のキャリア、または  
データ所有者  
あなたがやろうとしていること  
〜の結果としての画期的な研究  
アクセス権がある  
桁違いに詳しい情報  
それ以外の場合は他のデータサイエンティストよりも  
重要な先手となるだろうか  
あなたがそうであることがあなたにとって有利です  
早い  
このコースを受講することでこのように  
そして、あなたはある種の道徳も理解していると思います  
そして倫理的な  
肯定的かつ道徳的で倫理的な理由  
あなたは押し出すのを手伝っています  
この方向に押し進めてください  
学术界と政府  
コミュニティと民間部門  
コミュニティをこの方向に向けて  
対照的にこれらのテクノロジーを採用する  
これらの**一元化されたデータ**を作成するには、  
**データベース**を正しく理解する

9:52 authoritarian  
9:53 in nature and have significant risk of  
9:55 misuse that this is like a moral and  
9:56 ethical obligation of something that we  
9:58 must do so the the the you know these  
10:01 things are hand-in-hand they go together  
10:03 this should be a great way for you to  
10:05 build a career and have a competitive  
10:06 advantage of data science but also a  
10:08 moral and ethical outcome that you  
10:09 should be extremely proud to be a part  
10:11 of um and with that i hope you've really  
10:13 enjoyed this lesson and i hope you enjoy  
10:15 the next one and the rest of the course  
10:16 see you then

英語 (自動生成)

権威主義的な  
自然界では重大なリスクがある  
これが道徳のようなものであるという誤用  
私たちが行う何かの倫理的義務  
そうしなければなりません、あなたはこれらを知っています  
物事は連携して進んでいきます  
これはあなたにとって素晴らしい方法になるはずです  
キャリアを築き、競争力を持つ  
データサイエンスの利点だけでなく、  
あなたもたらす道徳的および倫理的な結果  
その一員であることを非常に誇りに思うべきです  
ええと、それでは、あなたが本当にそう思ってくれることを願っています  
このレッスンは楽しかったです、楽しんでいただければ幸いです  
次のコースと残りのコース  
それではまた会いましょう



OpenMined

<https://www.youtube.com/watch?v=r51F8rGIqfI>