

본스크립트는 KERI에서 운영하는 영어스크립트 제작봉사단(OES)의 영어재능기부를 통해 제작되었습니다.

* 번역자 : 강지연(성균관대), 백동진(서울과고), 한지혜(청심국제고), 이동현(고려대)

SDF2014

* 연사: 권혜진 / Kwon Hyejin

* 소속/직함: 한국탐사저널리즘센터 / 데이터저널리즘연구소장

Head of data journalism research center / Korea center for investigative journalism

* 연설: "진실"

"Truth"

(6:03)

네, 서울디지털포럼 네 번째 세션 ‘깨닫다’를 시작하도록 하겠습니다.

연설을 통해서 여러분들의 혁신적인 지혜도 깨닫기를 바라겠습니다.

**Okay, we will now start our fourth session
which focuses on ‘inspiration.’
I hope that you will also learn innovative wisdom
by listening to this speech.**

자, 그럼 첫 번째 연사를 소개해드리도록 하겠습니다.

So let me introduce our first speaker.

디지털 시대가 완벽하게 도입이 되면서 정보를 전달하는 매체들은 사실 정말 많아졌습니다.
권력을 제대로 감시하는 기능을 가지고 있는 매체들은 점점 더 줄고 있는 것도 사실입니다.

**As the digital era approaches,
the number of media that conveys information is increasing.
However, the number of media
that monitors power is decreasing.**

언론 본연의 기능을 잃어가는 미디어 매체의 현실과 해결 방안, 그리고 투명하고 보다 나은 세상을 만드는 데이터저널리즘의 세계로 여러분을 안내하도록 하겠습니다.

**You will now listen
how the media is currently losing its true function,
how to solve this problem and how to make
a transparent and better society through data journalism.**

한국탐사저널리즘센터 데이터저널리즘연구소장 권혜진 박사를 큰 박수로 맞아주시기 바랍니다.

**Please greet Kwon Hyejin from, who is
the head of the data journalism research center from
The Korea center for investigative journalism
with a big round of applause.**

안녕하세요. 뉴스타파에서 리서치디렉터와 데이터저널리즘연구소장을 맡고 있는 권혜진입니다.
반갑습니다.

**Hello. I am Kwon Hyejin.
I am the research director and head of the data journalism research center
at Newstapa.
It's a pleasure to meet you all.**

저는 오늘 진실과 데이터라는 키워드를 가지고 여러분들을 데이터저널리즘의 세계로 모시고자 합니다.

**I would like to invite you to the world of data journalism.
To do this, I brought two key words,
which are truth and data.**

저는 오랫동안 중앙일보, 동아일보를 포함해서 언론 현장에서 정보를 찾고 데이터를 분석하는 일들을 해왔는데요, 제가 대학에서 강의를 하기도 하는데, 뉴스타파라고 하는 작은 비영리 독립 탐사보도매체로 옮겨오게 된 것은 지금 서울디지털포럼의 주제, 그 주제와 관련이 있습니다.

**For a long time I have been finding information
and analyzing data in various media sites
including Joongang daily and donga daily.
I also did some lectures at universities.
Today's theme in Seoul Digital Forum is related to
my story of moving in to a
nonprofit independent investigative reporting media called Newstapa.**

‘혁신적 지혜, 기술에서 공공선을 찾다’라는 주제인데요, 저는 언론 분야에서 이 지혜를 데이터와, 그리고 탐사 저널리즘 속에서 찾고 있습니다.

**Our theme is about innovative wisdom
and technology for the common good.
I find this particular wisdom
in investigative journalism in the realms of journalism.**

언론에서 데이터가 중요한 것은 보다 많은 데이터가, 보다 풍부한 데이터가 시민의 알 권리를 확대하고, 또 진실을 향해 한 걸음 더 나아갈 수 있게 하기 때문인데요,

Data is important in the press because more data,

**meaning an abundant amount of data,
can not only lead to the citizens right to know,
but it can also help us move one step further to the truth.**

탐사보도가 중요한 것은 점점 미디어 산업의 경쟁이 격화되면서 이윤이라든가 상업적 동기에 의해서 권력 감시라고 하는 언론의 역할이 점차 약화되고 있다고 생각되기 때문입니다.

**Investigative reporting is important
because as the competition among media industries intensify
and companies seek for profits and commerce,
the media's role to watch over power abuse is weakening.**

언론의, 저널리즘의 위기를 느낀 해외의 많은 언론인들은 비영리 탐사보도매체들을 만들고 있고, 또 그것이 전 세계적으로 확대되는 추세죠.

**Many journalists from abroad
who feel that the true meaning of journalism has fade
made nonprofit investigative reporting media
and now it is spreading all over the world.**

탐사보도는 누군가가 감추고 싶어 하는, 그런 사회적으로 중요한 이슈를 발굴하고 폭로하는 일들을 합니다.

**Investigative reporting discovers and reveals
socially important issues that
some people want to hide.**

따라서 어느 누구에게는 불편한 진실이 될 수도 있죠.

So to a few people this can come as an inconvenient truth.

제가 오늘 여러분들한테 소개하는 것들이 또한 그렇게 될 수도 있을 거라고 생각이 듭니다.

**I believe some issues that I will talk about today
might come as an inconvenient truth to some of you as well.**

자, 저는 권력 감시, 그리고 조세 정의, 또 고위공직자와 시민의 알 권리, 이런 문제를 가지고 여러분들을 데이터저널리즘의 세계로 소개하고자 합니다.

**So I would like to introduce you to
the world of data journalism
by talking about observation of power, justice in taxation,
senior government officials, and people's right to know.**

2012년 12월 11일, 소위 국정원 직원으로 밝혀진 한 사람이 대통령선거와 관련된 글을 인터넷 사이트에 올리다가 발각이 되는 일이 있었습니다.

**In 2012, December 11th
a person who was later discovered
as a national intelligence service was found
writing a post related to the presidential election on the internet.**

여러 매체에 의해서 그 내용은 확인이 되었죠.
저희 취재진은 이런 합리적 의심을 했습니다.

**Several media checked this truth.
However, my fellow reporters made a reasonable suspicion.**

인터넷 사이트에 뭔가 조직적으로 글을 올렸다면 그것이 ‘오늘의 유머’라든가 ‘보배드림’같은 사
이트에 한정된 것이 아니라 트위터라든가 다른 곳에서도 있었을 것이라는 거죠.

**If an organization posted something the Internet,
then it would not be limited to sites such as
‘todayhumor’ or ‘bobaedream’
but also on twitter and other sites.**

그래서 알려진 내용을 인터넷 검색, 그리고 트위터 검색을 통해 확인해 봤습니다.

**So we searched the Internet and used twitter
based on the facts we knew.**

보시는 것처럼 내용이 거의 같은 것들이 발견이 되고, 그런 내용을 올린 아이디들이 발견이 되었
죠.

**So as you can see, we found a lot of similar posts
and discovered the user ID who posted it.**

그러면 그 계정을 대상으로 해서, 트위터를 하시는 분들은 아시겠지만, 팔로잉, 팔로워, 또 리트윗
아이디들을 수집해나가기 시작합니다.

**Starting from that twitter account,
those who use twitter may know but,
we could collect other IDs
who were following or retweeting that person.**

그 많은 아이디들 중에서 저희가 주목한 것은, 일정한 기준을 가지고 수집을 해 들어가기 시작했
는데, 그 기준은 이렇습니다.

**Among those IDs, we had a standard criteria
that we focused on to collect more information.
This was our standard.**

12월 11일을 기점으로 해서, 트위터 계정 활동이 정지되거나, 혹은 삭제한 아이디들, 그리고 특정 시점을 기준으로 해서 트위터 활동이 활발해졌는데 소셜 미디어 톨로 봤을 때 그것이 패턴이 비슷하게 나오는 것들, 내용적으로 봤을 때에는 중복 비판이라든가, 정책 홍보, 야당 비판과 같은 내용이 들어있는 트윗 멘션들입니다.

We focused on those whose twitter accounts were stopped, or erased after December the 11th. And those whose twitter accounts were active after a specific point of time, but in fact was showing an overall similar active pattern when viewed via social media tools. And tweet mentions about criticizing North Korea, advertising policies, and criticizing the opposing party.

지금 보시고 있는 것은 소셜 미디어 분석 톨을 통해서 본 트윗 멘션들의 활동 그래프입니다.

The graph that you are seeing is the tweet mentions and activities that were investigated by social media investigation tools.

보시다시피, 저 시점은 2012년 8월 20일 부근인데요,

As you can see that time was around 2012 August 20th,

종료된 시점, 트윗 활동이 정지되고 계정이 삭제된 시점은 바로 12월 11일 정도가 됩니다.

and the end date when tweet activities have stopped and accounts have been erased is around December 11th.

8월 20일은 여당 대통령 후보가 결정된 날이죠.

August 20th is when the president candidate of the opposition party got nominated.

저희는 이렇게 현장 취재가 아니라 데이터 취재를 통해서 진실들을 추적해가기 시작했습니다.

We covered these news not by collecting data at the scene, but by tracking these facts online.

그렇게 해서 저희가 수집한 아이디들은 약 600여 개가 넘었는데요,

As a result, we gathered around 600 tweet IDs,

그 600여 개가 올린 트위터 멘션들을 수집을 해야 되는데 거기에서 조금 난관에 부딪혔습니다.

and as we tried to gather the tweet mentions of these 600 IDs, we got into a little bit of trouble.

왜냐하면 짐작하시다시피 국내에서 그것을 줄 수 있는 데가 많지 않았거든요.

The reason is because, as some of you guys know,

**there are not many places
where they give you the information in Korea.**

그래서 저희는 별도로 인터넷 프로그램을, 크롤러 프로그래밍을 만들어서 해외에서 데이터를 수집했습니다.

**So we made our independent internet program
called crolor programming to gather information from abroad.**

그렇게 해서 수집한 트윗 데이터가 한 28만여 건이 됐는데요,

We were able to gather around 280,000 tweeter mentions,

그 28만여 건의 트윗 데이터를 가지고 소위 소셜 네트워크 분석을 했습니다.

**and with these tweeter mentions
we enrolled in a social network investigation.**

사회 연결망 소프트웨어를 이용했는데, 예를 들면 '게피' 라든지, '넷마이너' 같은 프로그램이죠.

**We used society connecting software
such as, 'gepi' or 'net minor'.**

그렇게 해서 소프트웨어로 분석을 해보면 리트윗 네트워크를 통해서 거기에서 연관관계가 드러나게 됩니다.

**So if we analyze using the software,
we can find the relationship through retweet networks.**

저희는 모듈러리티 분석이라는 것을 통해서 10개 그룹이 활동을 했다라는 분석을 내놓았죠.

**We ended up concluding that 10 groups were active,
using the modularity analysis.**

그리고 그 연루 의혹이 있는 아이디들이 올린 멘션을 들여다보면 대통령 선거와 관련된 내용이거나, 혹은 정치 관련된 내용들이 직접 올린 글은 한 3700여 건, 그리고 실제 리트윗한 것들은 55000건 이상이 발견되는 것을 볼 수 있었습니다.

**Also if we look into the mentions of the related IDs,
3700 cases are related to presidential election or other political issues,
and 55000 cases were actually retweeted.**

자, 같이 한 번 보시죠.

Now, lets take a look over here.

영상:

그 결과 얻어낸 네트워크 지도입니다.

This is the network map as a result of our analysis.

국정원 추정 계정을 포함해 모두 5500여개 계정이 서로 연결돼 있습니다.

**Including the speculated national intelligence service accounts,
5500 accounts are related to each other.**

초록색이 일반 계정이고, 빨간색이 국정원 추정 계정입니다.

**The green accounts are the ordinary ones
and the red accounts are the speculated
national intelligence service accounts.**

국정원 추정 계정만 남겨놓고 모두 지워봤습니다.

We erased all the other accounts except the red ones.

RT 횟수가 10번 미만인 계정도 연결도가 낮다고 판단해 지웠습니다.

**We also erased the accounts with less than
10 RT since we perceived that there is a low relationship.**

그랬더니 국정원 추정 그룹의 전체적인 관계망이 모습을 드러냅니다.

**Then suddenly, the overall relational net
of the red accounts show up.**

여기에 분석 과정에서 새로 확인된 의심 계정을 추가해, 네트워크 지도를 완성했습니다.

**Furthermore, adding some other suspicious accounts
that we found during the analysis,
we completed our network map.**

큰 원으로 표시된 핵심 계정이 주로 인용이 이루어진, 다시 말해 주요 콘텐츠 생산자입니다.

**The account marked by
the big circle is the core account,
the contents supplier.**

핵심 계정이 왕성하게 글을 작성하면 주변 계정들이 RT로 퍼 날랐음을 보여줍니다.

**The account marked by the big circle
is the core account, the contents supplier.**

권혜진:

저희가 발견해낸 아이디하고, 검찰 측에서 발견한 아이디들은 70% 이상이 비슷한 것으로 발견이
됐습니다.

The IDs that we found and those that the police found,

were said to be over 70% similar.

제가 다음에 소개드릴 내용은 조세 정의에 관한 프로젝트입니다.

The next thing that I want to introduce to you is the taxation justice project.

아마 여기에 계신 많은 분들이 작년에 이 보도를 접하셨을 텐데요,

Maybe a lot of you have seen this issue last year,

저희가 ICIJ, 즉 국제탐사언론인협회와 한국에서는 유일한 파트너로 참가해서 조세피난처 프로젝트를 진행했습니다.

we the ICIJ and Korea participated as the only partners in the tax haven project.

근데 이 조세피난처 프로젝트는 아주 전형적인 데이터 프로젝트였어요.

However, this project was a typical data project.

조세피난처, ICIJ 측에서 확보하고 있는 데이터는 파일로 따지면 200만개 이상이 넘는, 그리고 분량으로 따지면 260GB가 넘는 데이터를 갖고 있습니다.

The amount of data that ICIJ has related to the tax haven is more than 2,000,000 files or otherwise more than 260GB.

그 데이터 안에는 이메일이라든지, 이메일에 첨부된 파일이라든지, 아니면 pdf파일, 아니면 이미지, 워드 파일, 이런 것들이 한꺼번에 섞여있는 비정형 데이터였어요.

In that data there are e-mails or attached files, pdf files, image files, word files and other all sorts of atypical data.

그러한 비정형 데이터들을 전 세계적으로 참여하고 있는 데이터 저널리스트들은 데이터를 조직화하고 네트워크화하고, 또 이미지 파일들은 문자인식 프로그램을 통해서 검색이 되도록 하는 그런 작업을 거쳤습니다.

Data journalists all around the world have participated in organizing, structuring and networking unstructured data and image files with recognition program to swiftly restructure and search through characters in the data.

그렇게 해서 저희가 한국 측의 유일한 파트너로 참여하면서 매일같이 데이터를 뒤졌죠.

We participated as Korea's only partner and looked through the data every day.

왜냐하면 실명으로 확인된 어느 누구의 데이터를 준 것이 아니기 때문에 신원을 확인하고 취재를

하는 것은 온전히 저희의 몫이었습니다.

**Because we were not given the data of their identification,
it was our task to identify the person and to interview.**

저는 별도 사무실에서 계속 데이터를 들여다보는 일들을 했었는데요,

**I took the task of looking through
the continuously flowing data in another office,**

그렇게 밝혀진 조세피난처의 한국인이라는 프로젝트에서는 아시다시피 사회 명망가라든가 재벌 그룹들의 많은 사람들이 발표되기도 했습니다.

**Already stepped on tax haven,
many people of the conglomerate's reputation has been announced.**

전두환 전 대통령의 장남 전재국 씨 같은 경우에도 페이퍼 컴퍼니를 만든 것으로 드러났는데요,

**Chun Jae Guk, the eldest son of former president Chun Doo-hwan,
had a paper company.**

이 경우에는 그 페이퍼 컴퍼니가 한국에 주소지를 두고 있지는 않았지만, 여권 번호라든지, 아니면 그것과 관련된 다른 이메일에 첨부된 주소라든지 이런 것들 속에서 확인이 돼서 취재가 된 경우입니다.

**In this case, the paper company's address wasn't in Korea,
but passport number and addresses
attached to related emails were used to check
its existence and was soon reported.**

저희가 짧은 시간 동안 한국의 명단 속에서 밝힐 수 있는 것들은 굉장히 많이 밝혔지만, 그렇게 하지 못한 것들은 시민들의 참여를 통해서 취재를 하는 크라우드소싱 방식으로 취재를 진행해 나갔습니다.

**We found out a lot from the list in a short time,
but the rest we preceded using the 'crowd sourcing'
method and promoted public participation.**

왜냐하면 ICIJ가 전 세계의 10만 개가 넘는 페이퍼 컴퍼니 자료를 모두 공개할 것이라는 것을 저희는 알고 있었거든요.

**Because we already knew that ICIJ
will soon reveal data on over
one hundred thousand paper companies
all around the world.**

그렇기 때문에 한국에 주소지를 둔 사람들을 저희도 공개를 하고, 같이 발 맞춰서 공개를 하고, 많은 분들에게 제보를 해주십사 크라우드소싱 방식을 채택했습니다.

Therefore, we have adopted crowdsourcing to fit for the public whose address is in Korea, and who are willing to provide information with people.

이러한 시민참여형 프로젝트를 통해서 많은 제보가 들어왔고요, 그런 것들이 실제 보도에 반영되기도 했죠.

Such public participation projects provided many reports, and these reports were actually reflected in coverages.

장기간 걸쳐서 진행됐던 조세피난처 프로젝트는 올해 중국을 마지막으로 종결이 되었습니다.

This long term project on Tax Haven ended this year at China.

중국 프로젝트에 참여하면서 저희는 영국의 BBC라든가 가디언, 일본의 아사히 신문, 프랑스의 르몽드지와 같은 다른 뉴스 언론들과 같이 중국의 조세피난처 관련 보도들을 진행했습니다.

We also took part in the China project. We worked with BBC and The Guardian of UK, Asahi of Japan, Le Monde of France, and others, reporting on Chinese Tax Haven.

그러다보니 한 때 중국에서 ICIJ 홈페이지와 저희 뉴스타파 홈페이지가 차단되는 경우도 있었어요.

Come to think of it, ICIJ and News-Ta-Pa websites were once blocked.

관심 있는 분은 인터넷에 가셔서 'ICIJ Database' 라고만 구글에서 검색을 하시면 이런 사이트를 만나실 수 있습니다.

Those of you who are interested can google 'ICIJ Database' and you will see links to these sites.

지금 ICIJ에서 확보한 10만 개 이상이 넘는 페이퍼 컴퍼니에 대한 자료가 주소지까지 모두 전 세계에 공개가 되어 있어요.

Right now, ICIJ's data on over a hundred thousand paper companies is open to the world. Even their addresses.

그러다보니 저기에서 '부산' 이렇게 검색을 한다면, 사람 이름을 검색을 한다면, 이러면 주소의 번지수까지 나와 있는 경우들이 많이 있습니다.

So if you search 'Busan' or someone's name, most of the time, you will even get house numbers.

이 사이트를 만든 곳은 코스타리카 언론이에요.

Costa Rican media made this website.

그래서 전 세계 각 나라에 흩어져 있는 탐사 저널리스트들이, 특히나 데이터 저널리스트들이 이러한 전 세계의 공공선을 위해서 참여하고, 프로젝트를 진행해가고 있는 상황입니다.

Investigative journalists, especially data journalists around the world are working for the common good, carrying on various projects.

작년에 한국 사회를 흔들었던 조세피난처 보도는 저렇게 보시다시피 관세청의 보도 자료로도 확인할 수 있는데요,

**Tax Haven reports were big issues last year.
And we can see that in reports from the Customs Office.**

저희 보도에 의해서 “국부유출 등 불법외환거래 약 1조 123억 상당을 적발했다”고 발표를 했습니다.

**We announced the report
‘Caught Draining National Wealth
and Illegally Dealing Foreign Currency,
Over 1 Billion 123 million Won.’**

그 중에 보시면 “‘뉴스타파’가 공개한 명단과 관련해서, 신원을 확인 중이고, 우선 13개 업체의 재산도피 등 불법외환거래 총 7389억원 상당을 적발했다”라는 발표를 했습니다.

**In the report, it says
“We are still verifying identities on the list.
But we did expose 13 businesses for flying assets
abroad and illegal foreign exchange transaction.
It amounts to 7389 million won in sum.”**

한편으로는 저희가 언론으로서 입법, 사법, 행정에 이어서 제 4부라는 얘기를 하는데요,

**Sometimes, we commonly say press is the fourth branch,
after legislative, executive, and judicial branch.**

언론이, 그러한 건전한 감시의 역할을, 감시견의 역할을 한 것이 아닌가, 이런 생각이 듭니다.

**And I question myself.
Did the press play a role in
healthy surveillance and function as a watch dog?**

저희가 중점적으로 하고 있는 것 중에서는 예산에 관련된 것들을 중요하게 생각하고 있습니다.

We mainly focus on works related to budgets.

내가 세금을 내는데 그 세금이 어디에 쓰이는지, 혹은 그 세금이 쓸데없이 낭비되고 있는 것은 아닌지에 대한 것은 굉장히 중요한 문제라고 생각이 드는데요,

**I say where my tax goes, how it is used,
and whether it is wasted are all very important.**

사실 전문가가 아니면 그 예산 자료를 입수하고 분석하고 하는 것이 참으로 어렵습니다.

If you are not an expert,

**it's indeed hard to gather information
on budget and analyze them.**

어려워서, 저희 역시 어려움을 많이 겪고 있죠.

We, too, are going through much difficulty.

그래서 한 해에 끝나는 문제가 아니고, 이것을 장기 기획으로, 지속적으로 감시를 하는 프로젝트를 진행하고 있는데요, 그렇게 해서 많은 것들이 또 발표가 되고 있습니다.

**That's why we don't do yearly projects.
Instead, we plan on a long term basis and constantly keep watch.
This is how we announce most of the reports.**

예를 들어 최근에는 지자체에서 지급되고 있는 사회 보조금의 경우에 관변단체 보조금이 한 해에 350억 정도가 특정 단체들에게 지급됐다, 이런 것들이 밝혀지고 있는데,

**For example, recently, it has been unveiled
that in the case of social aid provided by local administrations,
interest group aid of 350 million won
was given to certain groups in one single year.**

그런 것은 저희가 누구에게 보도 자료를 받아서 쓰는 것이 아니라 직접 정보공개 청구를 한다든가 아니면 예산 관련된 자료들을 막 뒤진다든가 이런 것을 통해서 발굴해낸 것이죠.

**We didn't receive the information.
We either claim for release of the information
or search anything related to budgets.
This is how we make discoveries.**

작년에 서울디지털포럼에서 사이먼 로저스라고 하는 가디언지의 데이터 에디터 출신이 와서 발표를 한 적이 있습니다.

**Last year, Simon Rogers, data editor at The Guardian,
came to Seoul Digital Forum to talk.**

그 내용에 보면 그렇게 정보 공개가 잘 되어있는 영국 정부에서도 이 예산 데이터를 본인들이 하기 위해서는 수많은 pdf 자료를 뒤져서 만들어야 하기 때문에 어려움을 겪고 있다는 얘기를 했는데,

**In his talk, he also mentions that
even in a country like UK where information are fairly open,
it is very exhausting to search through all the PDF files.**

비단 우리나라의 상황만은 아니고 굉장히 많은 정부가 발표하는 자료들이 실제 시민들이 보기에 어려운 형태로 제공되고 있는 경우가 많이 있죠.

**This shows that not only the Korean government,
but also governments of other countries have**

their information in formats that are difficult to read.

저희가 지방선거를 앞두고 있는데요,

We have a provincial election coming up.

대통령선거라든가 지방선거를 앞두고 후보들이 공약을 내놓는데 우리는 몇 년이 지나면 어떤 공약을 내놓았는지 까먹기 일쑤입니다.

**When presidential or local elections are near,
candidates come up with pledges.**

But the thing is we all forget them after a couple of years.

그래서 그 공약이 얼마나 지켜졌냐는 이 공약의 이행 정도를 시민들이 알고 선거에서 '어떤 후보를 뽑아야 되겠다' 라는 정보를 줄 수 있어야 되는데, 그런 장기 기획 역시 하고 있습니다.

**So we are planning on a project to provide
information on how much each pledge was
actually put to execution.**

그래서 좀 꼼꼼하게, 예를 들어서 대통령 공약의 경우에도 201개의 공약으로 나눠서 꼼꼼하게 자료를 정리하고, 그것의 이행정도를 지속적으로 살펴보는 프로젝트를 하고 있습니다.

**This project carefully inspects
the 201 pledges of the president
and organizes classified data.**

지금 나온 화면은 미국의 '텍사스 트리뷴'이라고 하는 저희 같은 비영리 독립매체인데요,

**On the screen is the United States'
independent non-profit media called "Texas Tribune",**

거기에서 보시면 고위공직자들의 재산을 저렇게 볼 수 있게 하는데,

**If you look from there, you can see
the properties of the government officials,**

어떤 것이냐 하면, 법안을 만들고 집행하는 고위공직자들이 어떤 재산을 갖고 있는지에 따라서 혹시 이해충돌을 빚는 부분은 없는지를 시민들이 알아야 된다는 것이죠.

**Citizens have right to know what kind of properties
that enforcing high-level officials, those who create a bill, have
in order to understand the conflictual interest
that might occur with their personal interest and characteristics.**

보시면, 사람을 클릭해서 전부 다 재산을 볼 수 있도록 시민들에게 공개를 하고 있습니다. 이런 작업을 저희도 하고 있죠.

**As you can see, we have made it open for the public
so that people can see all of the properties held by clicking the person.**

우리나라에서 고위공직자들의 재산이 공개된 것은 그렇게 오래된 것은 아닙니다.

Publishing properties of government officials in Korea is not so old.

하지만 초반부에 공개된 자료들은 소위 확장자만 pdf고 내용은 이미지에요, 그냥 사진이거든요.

Few of the published extension materials in the opening were pdf, but the contents were image files. A photo.

그러면 그것을 문자인식 프로그램을 돌리고, 또 텍스트 데이터베이스로 만들고 하는 것들은 매우 지난한, 지루한 작업이 되고 있습니다.

Running character recognition program and creating a text database is a tedious work.

예를 들어서 국회의원들의 재산 내역은 국회 사이트 중에서도 국회 공보라는 데에 공개가 됩니다.

For example, among the site of the National Assembly, National Assembly publication breakdown and open the properties of government officials.

매년 공개가 되죠. 그런데, 이 사이트는 특이하게도 밖에서 구글 같은 사이트로 검색을 하시면 내부 pdf의 문자열이 잘 검색이 안 됩니다.

Although Published every year, the search engines like google can not search through the strings of internal pdf.

그러면 저 안에 서치, 검색박스에서 검색하면 잘 검색이 될까요? 검색이 잘 안 됩니다.

Then if you search in the search box, do we get the proper result of the search? Not a chance.

아는 국회의원 이름으로도 검색이 안 되는데요.

We can't even search through the names of the legislators we know.

제목에 있으면 검색이 되겠죠.

Then we can search for the title of the document.

하지만 보시다시피 1호, 2호, 3호, 4호 이렇게 올라와 있기 때문에 일반인들이 저 자료를 찾기는 어렵습니다.

However, as you can see, the documents unloaded are named No. 1, No. 2, No. 3, No. 4 thus making the people difficulty searching through the data.

저희는 알죠, 42호에 있거든요.

We know it's in No. 42

그렇게 입수를 한다 하더라도 또 벽에 부딪칩니다.

Even if we get the information, we hit another wall.

이 자료는 pdf 형태로 올라와 있는데, 예를 들어서, 국회의원들의 재산 자료를 다 합쳐서 시민은 우리 지역의 국회의원이 어떤 주식을 가장 많이 갖고 있는지, 혹은 전체 국회의원들을 다 토털 분석해보니 가장 많이 갖고 있는 주식은 뭔지 보고 싶은데, pdf 형태로 있다면 작업이 어렵죠.

**The data is in pdf format, for example,
if the citizens want to see the parliamentarians' property data
and see what stocks, by statistic, the entire parliamentarians of our area
is mostly holding, or see most stocks possessed in total to be analyzed,
it strains the work if it is in a pdf format.**

보시다시피 은행이나 증권은 하나의 셀에 다 몰아져 있기도 하고요.

그래서 저런 것들을 궁금해 하는 시민들은 있겠지만 데이터가 없기 때문에 보도들이 나올 수가 없습니다.

**So there will be citizens who might be curious about it,
but because there is no data, media cannot make the report.**

그래서 뉴스타파 같은 경우에는 저희, 특히 데이터 저널리즘 연구소는 현재 없는, 그러나 시민의 공공선, 공익과 관련된 주요한 데이터들을 우선 만들어야 되겠다고 생각을 하고, 저런 작업들을 하고 있습니다.

**Though we have no data Journalism Institute currently in our station,
with sole idea of public interest and principal data
relating to common good of the citizens, we have been working.**

그런 작업은 상당히 지루한 작업인데요, 예를 들어서 4대강과 관련된 예산을 우리가 보겠다, 그 예산을 보는데 4대강에 참여한 업체 전체를, 큰 박스 7개 정도 되는 종이문서를, 저희는 입수해서 그것을 들춰다 보면서 필요한 부분을 입력합니다.

**Such work is a very tedious process.
For example, if we want to see the budget
that went into the development of four major rivers,
we would have to gather entire enterprises' involvement
that have participated in the project,
which sums to about seven boxes of paper documents.
And then we would have to manually filter and enter
necessary information in order to see.**

그것과, 또 매년 우리가 입수할 수 있는 고액 후원금 데이터, 국회의원에 대한 고액 후원금 데이터를 크로스체크를 하죠.

Then we cross-check reports of expensive donation data that we get every year, and the high donation data of parliamentarians.

그렇게 해서 아주 많은 시간을 들인 끝에 4대강 공사에 참여한 업체들의 고액 정치후원금, 이런 보도를 하게 됩니다.

End of long period of searching and organizing, reports of high political contributonal involvement with companies that participated in four major river project come out.

그리고 저것은 '타블로'라고 하는 소프트웨어를 통해서 시각화를 한 것이죠.

That's a visualization of the software called "tablo".

아까 말씀 드린 것처럼 고위공직자의 재산은 이해충돌 관계를 봐야 되기 때문에 늘 중요하고,

As I said earlier, to understand the conflictual interest with the relations with the property of government officials, it is always important.

또 정보가 바뀌면 그런 재산관계 데이터들을 저렇게 인터랙티브를 만들어서 공개를 하곤 합니다.

Also with changing information, it creates 'interactive' relations with characteristics of property related data.

제가 저희가 했던 작업들을 꼭 소개를 드렸는데요.

I have introduced the works we have had.

지금 보시면 '데이터 저널리즘은 미래다', 인터넷 창시자인 팀 버너스-리가 얘기를 했는데요.

Internet founder, Tim Berners-Lee once told 'Data Journalism is the future'

이러한 데이터에 기반한 보도들은 좀 더 우리에게 언론의 본질적인 역할들, 예를 들어 아까 말씀 드린 정부 감시라든가, 감시견으로서의 역할을 훨씬 잘할 수 있도록 도와주는 강력한 무기가 됩니다.

These data structured coverage becomes a powerful weapon as these basis allows us to function as a media and to help monitor the government as we take the role of watchdog.

하지만 그것이 저절로 이루어지진 않겠죠. 그것을 할 수 있는 사람이 있어야 되는데 지금 한국 언론에서는 뉴스를 이해하는 개발자가 없고 개발을 이해하는 저널리스트가 없는 상황입니다.

However, that won't happen automatically. There is no developer that understand the news and there is no journalist that understand

he development in the Korean media.

별로 없죠. 그래서 많은 분들이, 테크놀로지에 강한 저널리스트들이 많이 이 분야에 들어오셔서 데이터 저널리즘을 통해서 세상을 바꿔나갈 수 있기를 기대합니다.

Not much else.

**So I hope that many people,
those strong with technology enter this field of journalism
to change the world through data journalism.**

감사합니다.

Thank you.